BACILLALES INFLUENCE QUALITY AND SAFETY OF DAIRY PRODUCTS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Sarah Marie Beno

December 2017

BACILLALES INFLUENCE QUALITY AND SAFETY OF DAIRY PRODUCTS

Sarah Marie Beno, Ph. D.

Cornell University 2017

Bacillales, an order of Gram-positive bacteria, are commonly isolated from dairy foods and at various points along the dairy value chain. Three families of Bacillales are analyzed in this work: (i) Listeriaceae (represented by *Listeria monocytogenes*), (ii) Paenibacillaceae (represented by *Paenibacillus*), and (iii) Bacillaceae (represented by the *Bacillus cereus* group). These families impact both food safety and food quality. Most Listeriaceae are non-pathogenic, but *L. monocytogenes* has one of the highest mortality rates of foodborne pathogens. *Listeria* spp. are often reported in food processing environments. Here, 4,430 environmental samples were collected from 9 small cheese-processing facilities and tested for *Listeria* and *L. monocytogenes*. Prevalence varied by processing facility, but across all facilities, 6.03 and 1.35% of samples were positive for *L. monocytogenes* and other *Listeria* spp., respectively. Each of these families contains strains capable of growth at refrigeration temperatures. To more broadly understand milk spoilage bacteria, genetic analyses were performed on 28 *Paenibacillus* and 23 *B. cereus* group isolates. While no specific genes were significantly associated with cold-growing *Paenibacillus*, the growth variation and vast genetic data introduced in this study provide a strong foundation for the development of detection strategies. Some species within the *B. cereus* group have previously been shown to grow at refrigeration temperatures, but the genetic analyses provided here will be of importance for the development of screening tools and to more successfully assess spoilage risk. Overall, the work presented here covers groups of bacteria that are common challenges for the dairy industry and provides large data sets to encourage additional research surrounding these topics.

BIOGRAPHICAL SKETCH

Sarah Marie Beno was born on January 17, 1991 in Marietta, Georgia as the oldest of three children to Paul and Kathleen Beno. Sarah attended Meredith College from 2009-2013 and graduated from the Honors Program with a B.S. in Biological Sciences and a B.A. in Chemistry. At Meredith College, Sarah participated in undergraduate research, where she studied the African legume, *Tylosema esculentum*. Throughout her coursework, she gained an interest in microbiology and decided to pursue food microbiology in graduate school. Sarah began a Ph.D. program in Food Science, Microbiology, and International Development following graduation. Her committee, Drs. Martin Wiedmann, Randy Worobo, and Ralph Christy, provided many opportunities to connect with industry members and to apply this knowledge internationally. Sarah spent three months working with the International Potato Center (Centro Internacional de la Papa; CIP) in Nairobi, Kenya to develop a food safety training program and laboratory that can be used in junction with the sweetpotato value chain. Other international experiences relevant to her doctoral work include a two-week trip looking at value addition in food processing throughout India, and a trip to Uppsala, Sweden to participate in the Global Challenges University Alliance summer program focused on food quality, food safety, and food security. Sarah will next begin an IRACDA-MERIT fellowship program at the University of Alabama-Birmingham where she will train in both biomedical research and teaching, in preparation for a career in academia. Outside of academia, Sarah enjoys horseback riding, volunteering at the local SPCA (from where she adopted her fluffy companion, Lister), and experiencing new places.

To my family and friends for all of their love and support throughout this journey.

ACKNOWLEDGMENTS

I would like to thank Martin Wiedmann, my committee chair, advisor, and the principal investigator of the research presented here. Martin is challenging, yet generous, and I have benefitted extraordinarily from the lessons, support, and network that Martin shares. I also thank Kathryn Boor for always providing advice and encouragement. Kathryn is a true role model- hard working, successful, and kind. I would also like to thank my committee members, Randy Worobo and Ralph Christy, for their time, support, and useful connections (especially for my international trips).

I owe many teaching and presenting opportunities, trips, and Taco Tuesdays to the Cornell Dairy Extension Team. The Dairy Extension Team has been instrumental in gaining teaching and extension opportunities throughout graduate school. I would especially like to thank Rob for helping me get started with the artisan cheese project, Kim for traveling with me frequently and teaching me about the world of auditing, and Louise for always seeking to improve our program. You have all been wonderful friends to me. Furthermore, the FSL/MQIP group has been a pleasure to work with. I would like to thank Maureen Gunderson for always honoring media requests and keeping the lab going. Thank you all for your friendship and for always working together. This lab is truly something special. Go Wiedmann Booriors!

Of course, I must thank my family and my friends for all of the support. Thank you to my parents for always encouraging me and helping me get here! To Conor, for enduring another four years of long distance and for always offering your support. To my favorite housemates- Rachel Evanowski, Mary Kate Koch, Sam Hempel, Courtney Smith, and Claire Burtch, for bringing me laughter every day. Thank you to my friends (both human and furry) at If Only Farm and the Tompkins County SPCA for the happiness and love. A special thanks to my favorite fluffy gray friends, Jiminy and Lister, for always making me smile.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Bacillales, an order of Gram-positive bacteria, is commonly associated with concerns for both food safety and food quality. These organisms include *Listeria monocytogenes*, the *Bacillus cereus* group, and *Paenibacillus*. *Paenibacillus* and *Bacillus* are the two most influential genera for Gram-positive dairy spoilage (Ivy et al., 2012) and *L. monocytogenes* is a well-characterized foodborne pathogen. These three genera are important challenges for the dairy industry.

In 2012, Ivy et al. characterized 1,288 spore-formers from dairy-associated sources and found that 68.7% of the *Paenibacillus* isolates represented the species *P. odorifer*. While *P. odorifer* is often able to grow at refrigeration temperature and can cause off-flavors and curdling in fluid milk (De Jonghe et al., 2010; Ranieri et al., 2012), the *B. cereus* group interestingly has ties to both food safety and food quality; certain members of the *B. cereus* group are associated with spoilage. Some strains of the *B. cereus* group are able to grow at low temperatures (Lechner et al., 1998; Miller et al., 2016), just like *Paenibacillus* (Moreno Switt et al., 2014) and *L. monocytogenes* (Chan and Wiedmann, 2009), while others have toxins that have led to diarrhetic and emetic disease (Bottone, 2010). Illnesses caused by bacteria in the *Bacillus cereus* group are typically short-lived and do not result in serious long-term injury (Ehling-Schulz et al., 2004). On the other hand, *L. monocytogenes* has one of the highest mortality rates of all foodborne disease (some estimates as high as 24%), even though it is opportunistic and generally only affects a small population (Farber and Peterkin, 1991).

Fluid milk, cheese, and ice cream have all been associated with foodborne illness in recent years (Chen et al., 2017a; Chen et al., 2017b; Costard et al., 2017) These outbreaks were sourced to both raw and pasteurized products, suggesting that even with pasteurization

as a kill step, products are still high-risk. Ready-to-eat food products, such as cheeses, are subject to environmental contamination (Tompkin, 2002). Environmental pathogen monitoring programs can be used to verify critical control points and to eliminate harborage sites for bacteria throughout the processing facility. Ubiquitous bacteria, such as *Listeria*, are frequently isolated from dairy processing environments (Leong et al., 2014) and may lead to contamination of food products, potentially resulting in foodborne illness.

Listeriosis, the disease that results from an infection with *L. monocytogenes*, most commonly affects a high-risk population comprised of the elderly, infants, immunocompromised individuals, and pregnant women. These infections can lead to septicemia, meningitis, and spontaneous abortion (Farber and Peterkin, 1991; CDC, 2017). The United States Centers for Disease Control and Prevention (CDC) estimates that 1,600 people become ill from *L. monocytogenes* each year and 260 die (16% motality). The CDC also reports that 90% of listeriosis infections are in the high-risk population (CDC, 2017). Of the 17 *Listeria* species, only *L. monocytogenes* is typically pathogenic in humans (Gray et al., 2004), but other species are used as index organisms. *Listeria* can persist in food processing environments and using the seek and destroy approach is valuable to ensure a facility is using proper precautions against *Listeria* and other foodborne pathogens (Tompkin, 2002).

Although not pathogenic, *Paenibacillus* can survive in a variety of stress conditions (e.g., low temperature, high mineral concentrations, etc.) and are able to enter the milk value chain at any stage from "grass to glass." Importantly, *Paenibacillus* is a spore-forming genus, and spores are known to survive pasteurization (Gopal et al., 2015). *Paenibacillus* spp. are frequently associated with fluid milk contamination and spoilage (Ivy et al., 2012; Ranieri et al., 2012; Moreno Switt et al., 2014). As *Paenibacillus*, primarily *P. odorifer*, can cause milk spoilage, these organisms may lead to quality issues including sensory defects (curdling) and decreased shelf life (Grady et al., 2016). As we aim to become more food secure, strategies against food spoilage are of increasing importance. The work presented here investigates the

links between a *Paenibacillus* genome and an isolate's ability to grow at refrigeration temperatures.

Interestingly, the *B. cereus* group comprises strains that cause spoilage and strains that cause foodborne illness. While *L. monocytogenes* rarely causes disease, members of the *B. cereus* group are responsible for at least 63,000 illnesses in the United States each year (Scallan et al., 2011). While few annual hospitalizations (20) and no deaths have occurred from foodborne illness caused by members of the *B. cereus* group, *B. cereus* has a large reach but often goes undiagnosed (Scallen et al., 2011). Organisms in this group can cause either emetic or diarrheal disease (Scallan et al., 2011). However, the impact of the *B. cereus* group goes beyond foodborne illness; the *B. cereus* group is often isolated from HTST fluid milk (Ivy et al., 2012) and has been shown to cause spoilage (Gopal et al., 2015). In fact, a 2006 study (Scheldeman et al.) found that psychrotolerant *B. cereus* group isolates were the most predominant raw milk contaminators during summer months. The genetic analyses presented here provide a large set of data with which to develop detection methods for the *B. cereus* group throughout the dairy product value chain.

Overall, food quality and food safety are engaging challenges across the food value chain. Farmers, the processing industry, and consumers are all stakeholders for these issues. The research presented here provides important information for the development of new processes to detect potentially problematic organisms in the dairy value chain, including the detection of pathogens like *L. monocytogenes* through environmental pathogen monitoring and a deeper understanding of cold growth genetics for spore-forming Bacillales influencing both product safety (the *B. cereus* group) and product quality (*Paenibacillus* and the *B. cereus* group).

REFERENCES

Bottone, E. J. 2010. *Bacillus cereus*, a volatile human pathogen. *Clin. Microbiol. Rev.* 23(2):382-398. http://dx.doi.org/10.1128/CMR.0073-09.

Centers for Disease Control and Prevention (CDC). 2017. *Listeria* (Listeriosis): People at Risk. http://www.cdc.gov/listeria/risk. Accessed on 17 July 2017.

Chan, Y. C. and M. Wiedmann. 2009. Physiology and genetics of *Listeria monocytogenes* survival and growth at cold temperatures. *Crit. Rev. Food Sci. Nutr*. 49(3):237-253. http://dx.doi.org/10.1080/10408390701856272.

Chen, Y., Y. Luo, H. Carleton, R. Timme, D. Melka, T. Muruvanda, C. Wang, G. Kastanis, L. S. Katz, L. Turner, A. Fritzinger, T. Moore, R. Stones, J. Blankenship, M. Salter, M. Parish, T. S. Hammack, P. S. Evans, C. L. Tarr, M. W. Allard, E. A. Strain, and E. W. Brown. 2017. Whole genome and core genome multilocus sequence typing and single nucleotide polymorphism analyses of *Listeria monocytogenes* associated with an outbreak linked to cheese, United States, 2013. *Appl. Environ. Microbiol*. http://dx.doi.org/10.1128/AEM.00633-17.

Chen, Y., Y. Luo, P. Curry, R. Timme, D. Melka, M. Doyle, M. Parish, T. S. Hammack, M. W. Allard, E. W. Brown, and E. A. Strain. 2017. Assessing the genome level diversity of *Listeria monocytogenes* from contaminated ice cream and environmental samples linked to a listeriosis outbreak in the United States. *PLoS ONE*. 12(2):e0171389. http://dx.doi.org/10.1371/journal.pone.0171389.

Costard, S., L. Espejo, H. Groenendaal, and F. J. Zagmutt. 2017. Outbreak-related disease burden associated with consumption of unpasteurized cow's milk and cheese, United States, 2009-2014. *Emerg. Infect. Diseases*. 23(6): 957-964. http://dx.doi.org/10.3201/eid2306.151603

De Jonghe, V., A. Coorevits, J. De Block, E. Van Coillie, K. Grijspeerdt, L. Herman, P. De Vos, and M. Heyndrickx. 2010. Toxinogenic and spoilage potential of aerobic spore-

formers isolated from raw milk. *Microbiol*. 136(3):318-325. http://dx.doi.org/10.1016/j.ijfoodmicro.2009.11.007.

Ehling-Schulz, M., M. Fricker, and S. Scherer. 2004. *Bacillus cereus*, the causative agent of an emetic type of food-borne illness. *Mol. Nutr. Food Res*. 48(7):479-487. http://dx.doi.org/10.1002/mnfr.200400055

Farber, J. M. and P. I. Peterkin. 1991. *Listeria monocytogenes*, a foodborne pathogen. *Microbiol. Rev*. 55(3):476-511.

Gopal, N., C. Hill, P. R. Ross, T. P. Beresford, M. A. Fenelon, and P. D. Cotter. 2015. The prevalence and control of *Bacillus* and related spore-forming bacteria in the dairy industry. *Front. Microbiol*. 6:1418. http://dx.doi.org/10.3389/fmicb.2015/0418.

Grady, E. N., J. MacDonald, L. Liu, A. Richman, and Z. Yuan. 2016. Current knowledge and perspectives of *Paenibacillus*: a review. *Microb. Cell Fact*. 15:203. http://dx.doi.org/10.1186/s12934-016-0603-7.

Gray, M. E., R. N. Zadoks, E. D. Fortes, B. Dogan, S. Cai, Y. Chen, V. N. Scott, D. E. Gombas, K. J. Boor, and M. Wiedmann. 2004. *Listeria monocytogenes* isolates from food and humans form distinct, but overlapping populations. *Appl. Environ. Microbiol.* 70(10):5833-5841. http://dx.doi.org/10.1128/AEM.70.10.5833-5841.2004.

Ivy, R. A., M. L. Ranieri, N. H. Martin, H. C. den Bakker, B. M. Xavier, M. Wiedmann, and K. J. Boor. 2012. Identification and characterization of psychrotolerant sporeformers associated with fluid milk production and processing. *Appl. Environ. Microbiol*. 78(6):1853-1864. http://dx.doi.org/10.1128/AEM.06535-11.

Lechner, S., R. Mayr, K. P. Francis, B. M. Pruβ, T. Kaplan, E. Wieβner-Gunkel, G. S. A. B. Stewart, and S. Scherer. 1998. *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. *Int. J. Sys. Bacteriol*. 48:1373-1392. http://dx.doi.org/10.1099/00207713-48-4-1373.

Leong, D., A. Alvarez- Ordóñez, and K. Jordan. 2014. Monitoring occurrence and persistence

of *Listeria monocytogenes* in foods and food processing environments in the Republic of Ireland. *Front. Microbiol*. 5:436. http://dx.doi.org/10.3389/fmicb.2014.00436.

Moreno Switt, A. I., A. D. Andrus, M. L. Ranieri, R. H. Orsi, R. Ivy, H. C. den Bakker, N. H. Martin, M. Wiedmann, and K. J. Boor. 2014. Genomic comparison of sporeforming bacilli isolated from milk. *BMC Genomics*. 15:26. http://dx.doi.org/10.1186/1471-2164-14-26.

Ranieri, M. L., R. A. Ivy, W. R. Mitchell, E. Call, S. N. Masiello, M. Wiedmann, and K. J. Boor. 2012. Real-time PCR detection of *Paenibacillus* spp. in raw milk to predict shelf life performance of pasteurized fluid milk products. *Appl. Environ. Microbiol*. 78(16):5855-5863. http://dx.doi.org/10.1128/AEM.01361-12.

Scallan, E., R. M. Hoekstra, F. J. Angulo, R. V. Tauxe, M. A. Widdowson, S. L. Roy, J. L. Jones, and P. M. Griffin. 2011. Foodborne illness acquired in the United States- major pathogens. *Emerg. Infect. Diseases*. 17:7-15.

Scheldeman, P., L. Herman, S. Foster, and M. Heyndrickx. 2006. *Bacillus sporothermodurans* and other highly heat-resistant spore formers in milk. *J. Appl. Microbiol*. 101(3):542-555. http://dx.doi.org/10.1111/j.1365-2672.2006.02964.x

Tompkin, R. B. 2002. Control of *Listeria monocytogenes* in the food-processing environment. *J. Food Prot*. 65(4):709-725.

CHAPTER 2

DEVELOPMENT AND VALIDATION OF PATHOGEN ENVIRONMENTAL MONITORING PROGRAMS FOR SMALL CHEESE PROCESSING FAICLITIES

Published in 2016, Journal of Food Protection 79:2095-2106

ABSTRACT

Pathogen environmental monitoring programs (EMPs) are essential for food processing facilities, of all sizes, producing Ready-To-Eat food products that are exposed to the processing environment. We thus developed, implemented, and evaluated EMPs targeting *Listeria* spp. and *Salmonella* in nine small cheese processing facilities, including seven farmstead facilities. Individual EMPs with monthly sample collection were designed specifically for each facility. *Salmonella* was detected in only one facility, with likely introduction from the adjacent farm, as supported by pulsed-field gel electrophoresis data. *Listeria* spp. were isolated from all nine facilities during routine sampling. The overall *Listeria* spp. and *L. monocytogenes* prevalences in the 4,430 environmental samples collected were 6.03% and 1.35%, respectively. Molecular characterization and subtyping data suggested persistence of a given *Listeria* spp. strain in seven facilities and *L. monocytogenes* persistence in four facilities. To assess routine sampling plans, validation sampling for *Listeria* spp. was performed in seven facilities following at least six months of routine sampling; this sampling was performed by independent individuals and included collection of 50 to 150 samples per facility, based on statistical sample size calculations; two of the facilities showed significantly higher *Listeria* spp. detection frequency in the validation sampling as compared to routine sampling, while two other facilities showed significantly lower detection frequency. This study provides a model for a science- and statistics- based approach to develop and validate pathogen environmental monitoring programs.

INTRODUCTION

Contamination of Ready-To-Eat (RTE) foods with bacterial pathogens is often attributed to cross-contamination that originates from the processing facility environment and occurs after the pathogen kill step (e.g., heat treatment). Cross-contamination from the processing facility environment has also been responsible for a number of foodborne illness outbreaks (Ferreira et al., 2014). While *Listeria monocytogenes* is a key foodborne pathogen that typically contaminates RTE foods from sources in the processing facility environments, other foodborne pathogens, such as *Salmonella*, can also contaminate RTE foods from the processing facility and other "post-kill step" environments. For example, two *Salmonella* outbreaks in 1998 and 2008, both linked to breakfast cereal produced in one specific processing facility, appeared to have been due to contamination from the processing facility's environment, with persistence of the causative *Salmonella* Agona strain in this facility for at least 10 years (Russo et al., 2013). *Salmonella* has also been found in dairy processing facilities, raw milk, and finished dairy products (Hedberg et al., 1992; Desenclos et al., 1993; Cody et al., 1999; Kousta et al., 2010; Carrasco et al., 2012). Post- pasteurization contamination of dairy products has also been linked to human salmonellosis outbreaks; for example, cross-contamination of pasteurized ice cream mix from a *Salmonella*- contaminated tanker truck was responsible for a large human salmonellosis outbreak in the U.S. in 1994 (Hennessy et al., 1996). Also, a multistate outbreak of two uncommon *Salmonella* serovars (Javiana and Oranienburg) in 1989 was linked to contaminated mozzarella from a single processing facility; contamination likely occurred in the processing facility, either from environmental sources or infected facility workers (Hedberg et al., 1992).

Globally, *L. monocytogenes* contamination of certain dairy products (e.g., surface ripened soft cheeses, fresh cheeses, washed-rind cheeses) represents a particular concern as a number of foodborne listeriosis outbreaks have been linked to these types of dairy products (Bille et al., 2007; Acciari et al., 2015; Heiman et al., 2015; Magalhães et al., 2015). In the

vast majority of cases, listeriosis outbreaks due to contaminated dairy products (including those made from pasteurized milk) have been linked to cross-contamination from the processing facility environment (Kousta et al., 2010; Gaulin et al., 2012; Gould et al., 2014; Melo et al., 2015). For example, a 2013 listeriosis outbreak with six hospitalizations, one miscarriage, and one death was linked to semi-soft French style cheeses produced in the U.S. (Food and Drug Administration, 2013). Pulsed-field gel electrophoresis (PFGE) showed that the subtype of *L. monocytogenes* isolated from the infected individuals matched isolates from contaminated cheese, as well as *L. monocytogenes* isolated from the environment of the processing facility two and three years prior to the outbreak, suggesting environmental contamination sources. Similarly, in a 2012 outbreak (twenty hospitalizations and four deaths) linked to a pasteurized ricotta cheese imported to the U.S. from Italy (Centers for Disease Control, 2010; Acciari et al., 2015), follow-up sampling revealed that *L. monocytogenes* isolates from the environment of the processing facilities had PFGE patterns that matched clinical isolates, suggesting that the cheese became contaminated in the processing facilities (Acciari et al., 2015).

The importance of the processing facility environment as a source of pathogens (especially *L. monocytogenes*) that cross-contaminate finished product has also been supported by studies that have investigated RTE foods and processing facilities without links to human disease outbreaks or cases (Ferreira et al., 2014). In many of these studies, molecular methods not only identified *L. monocytogenes* isolates with identical subtypes in finished products and the processing facility environment, but they also revealed that *L. monocytogenes* with identical subtypes were isolated over multiple sample collection dates across months and years (Almeida et al., 2013; Parisi et al., 2013; Barancelli et al., 2014; Stessl et al., 2014). For example, Lappi et al. (2004) reported re-isolation of the same *L. monocytogenes* subtypes in three out of four smoked seafood plants that were sampled over two years. Other studies have shown that *L. monocytogenes* can survive for several years in

dairy processing facilities (Unnerstad et al., 1996; Miettinen et al., 1999). Based on these types of findings, "persistence" of *L. monocytogenes* in processing facilities has been recognized as an important source of product contamination. Persistence is generally defined as a strain surviving in a facility over time (Carpentier and Cerf, 2011; Almeida et al., 2013). In many cases, isolation of a given subtype at three different sample collection dates has arbitrarily been defined as evidence for persistence (Kabuki et al., 2004; Carpentier and Cerf, 2011; Simmons et al., 2014).

Based on the importance of the processing facility environment as a source of *L. monocytogenes* and other foodborne pathogens (e.g., *Salmonella*), pathogen environmental monitoring programs (EMPs) represent an important tool that can help assess the effectiveness of food safety plans currently in place in a given food processing facility (Tompkins, 2002). Appropriately designed and implemented EMPs not only help detect locations where pathogens may survive (so called "niches") or be introduced but can also help validate and verify sanitation procedures and equipment and facility sanitary design. The importance of EMPs is also recognized in the *Preventive Controls for Human Foods* rule of the U.S. Food and Drug Administration (FDA) Food Safety Modernization Act (FSMA), which requires FDA regulated food processing facilities with RTE foods that are exposed to the environment to implement an environmental monitoring program (Food and Drug Administration, 2015).

Despite the recognized risk of *L. monocytogenes* contamination of some types of cheeses, there are limited field data on the development and implementation of EMPs targeting *Listeria* spp. (including *L. monocytogenes*) in small cheese processing facilities. In addition, the importance of environmental *Salmonella* sources in small and farmstead cheese operations is not well-defined. We thus developed, implemented, and evaluated EMPs targeting *Listeria* spp. and *Salmonella* in nine small cheese processing facilities. Subtyping of isolates was performed to gain a better understanding of pathogen transmission patterns and

persistence. This project also included a specific effort to develop and implement an approach to scientifically validate that EMPs and the associated sampling plans appropriately assess environmental pathogen contamination.

MATERIALS AND METHODS

Study design and facility selection. Nine small cheese processing facilities were selected based on willingness to participate, location, and frequency of production. According to FSMA, a small cheese processing facility is defined as one with 500 or less employees; all facilities in this study meet that classification (Food and Drug Administration, 2015). Five facilities (A, C, D, E, and F) were initially selected to start environmental sampling in April 2013 (Facility B dropped out of the study prior to the first sampling; Table 2.1). Two of these facilities (C and D) shut down before a whole year of sampling was completed. We thus selected four additional facilities (G, H, I, and J) to include in this study; sampling of these facilities started in May or June of 2014 (Table 2.1).

All facilities selected are located in the Northeast region of the U.S. and process cheese at least seven months of the year. Facilities A, E, G, H, and J used milk from two or three species (i.e., cows, goats, sheep) while the others produced cheese using milk from only one species. Seven of the nine facilities are classified as farmstead while two are classified as stand-alone (Table 2.1). The farmstead facilities are located within 100 m of the dairy farm that provides raw milk for processing. Farmstead facilities have been defined as facilities that process cheese on the same farm that provides raw milk (American Cheese Society). Several of these farmstead facilities share walls with animal housing.

TABLE 2.1: Summary of facilities sampled

Facility[a]	Environment[b]	Approx. production vol. (lbs milk/year)	Processing	Times sampled	Sampling Period	Environmental *Listeria* spp. (LS)[d] and *L. monocytogenes* (LM) sampling			Environmental, *Salmonella* sampling[e]	
						Total samples[f]	LS prevalence (%)	LM prevalence (%)	Total samples	Prevalence (%)
A	Farmstead	874,000	Year-round	24	April 2013-March 2015	917	29 (3.2)	12 (1.3)	156	0 (< 0.6)
C	Farmstead	30,000	Year-round	10	April 2013-January 2014	257	32 (12)	15 (5.8)	91	1 (1.1)
D	Farmstead	21,000	Seasonal	9	April 2013-May 2014	202	3 (1.5)	2 (0.99)	82	0 (< 1.2)
E	Farmstead	420,000	Year-round	24	April 2013-March 2015	1000	127 (12.7)	7 (0.7)	185	0 (< 0.5)
F	Farmstead	Not available[c]	Seasonal	14	April 2013-October 2014	422	4 (0.9)	0 (< 0.2)	74	0 (< 1.4)
G	Farmstead	860,000	Year-round	12	May 2014-April 2015	596	25 (4.2)	8 (1.3)	NS	NS
H	Stand-alone	500,000	Year-round	12	May 2014-April 2015	259	35 (13.6)	13 (5.02)	NS	NS
I	Stand-alone	35,000	Year-round	12	May 2014-April 2015	501	1 (0.2)	0 (< 0.2)	NS	NS
J	Farmstead	70,000	Year-round	12	June 2014-May 2015	276	11 (4.0)	3 (1.1)	NS	NS

[a] Facility B dropped out of the study prior to the first sampling and therefore, is not included in the facility list; facility I is the Cornell University cheese processing facility.

[b] Farmstead facilities are located within 100 m of a dairy farm that provides raw milk for processing

[c] The production volume for Facility F number was not provided; we estimated that this facility processes < 50,000 lbs of milk/year

[d] *Listeria* spp. numbers do include *L. monocytogenes* positive samples

[e] NS= not sampled

[f] This includes samples collected during routine and verification sampling

Sample sites and sampling procedure. Each facility was initially visited by at least two of the co-authors to develop individualized sampling plans. Sample sites were categorized by risk level according to the zoning concept as previously outlined (Grocery Manufacturers Association, 2014; Malley et al., 2015). Briefly, Zone 1 sites are product contact surfaces, Zone 2 sites are indirect contact surfaces that are physically close to the exposed product, and Zone 3 sites are those away from the exposed product but still in the general vicinity (e.g., drains, floors, and walls). Zone 4 sites are those outside of the exposed product area (Malley et al., 2015). In this study, we only sampled Zones 2-4, as sampling product contact surfaces would require fully-developed food safety plans, including improved traceability and hold programs. Examples of sites sampled in all or most facilities include drains, floors in the production area, and storage racks (a summary of all routine sampling sites is available in Supplemental Table 2.1, accessible from: https://foodsafety.foodscience.cornell.edu/research-and-publications/supplementary-materials-manuscripts/2016 and in Appendix A).

Environmental samples were collected monthly during processing using separate sterile sponges saturated with 10 ml of neutralizing buffer (3M, St. Paul, MN) for each site. Raw milk samples were also collected from processing facilities when available. Each facility shipped samples overnight to Cornell University's Food Safety Laboratory, and samples were processed upon arrival. Samples were tested for *Listeria* spp., including *L. monocytogenes*, as detailed below. In addition, environmental and raw milk samples from five facilities (A, C, D, E, and F) were collected every 3 months ("quarterly") over the initial 1-year period and tested for *Salmonella;* the same environmental sites sampled for *Listeria* spp. were sampled for *Salmonella.*

Sample enrichment, plating, isolation, and confirmation of *Salmonella*. Testing for *Salmonella* was performed according to the FDA Bacterial Analytical Manual (BAM) protocol (Andrews et al., 2014) with minor modifications as detailed below. Briefly, sponge

samples were added to 225 ml Lactose Broth (Becton, Dickinson and Company [BD], Sparks, MD), followed by manual homogenization. For raw milk samples, 25 ml of raw milk was added to 225 ml of Lactose Broth and sample bags were manually homogenized. After initial incubation at room temperature for 4 h, enrichments were incubated at 35 ± 2°C for an additional 24 h. After this primary enrichment step, 100 µl and 1 ml aliquots of the enrichment were added to 9 ml of Rappaport-Vassiliadis (RV; BD) and Tetrathionate (TT; BD) broth, respectively. After 24 h of incubation in a 42 ± 2°C shaking water bath, 50 µl of each of these secondary enrichments were plated onto CHROMagar *Salmonella* (CHROMagar, Paris, France) and Xylose Lysine Deoxycholate (XLD; BD) agar, followed by incubation at 35 ± 2°C for 24 h. On CHROMagar, typical *Salmonella* appear as mauve colonies, while on XLD agar typical *Salmonella* appear as red colonies, usually with black centers. For each sample, up to four colonies that showed typical growth and morphology on CHROMagar or XLD were sub-streaked to CHROMagar, which was subsequently incubated for 24 h at 35 ± 2°C. Typical *Salmonella* colonies were sub-streaked onto BHI agar, which was incubated for 24 h at 37 ± 2°C. Up to four putative *Salmonella* colonies per sample were then tested with a PCR assay targeting *invA*, which allows for identification of *Salmonella (38)*. A sample that yielded typical *Salmonella* colonies that tested positive with this PCR assay was classified as a "confirmed *Salmonella* positive". *Salmonella* isolates were stored in 15% glycerol at -80°C and used for further analysis with PFGE.

Sample enrichment, plating, isolation, and confirmation of *Listeria* spp. and *L. monocytogenes*. Samples were enriched according to a modified FDA BAM protocol (Hitchins et al., 2016). Briefly, 90 ml of buffered *Listeria* enrichment broth (BLEB; BD) was added to each sponge, followed by homogenization at 260 rpm for 60 s in a Stomacher 400 Circulator (Seward Ltd., United Kingdom). For raw milk samples, 25 ml of milk was added to 225 ml of BLEB and homogenized by hand. After enrichments were incubated at 30 ± 2°C for

an initial 4 h, 4 µl of *Listeria* selective enrichment supplement (LSES; Oxoid, Cambridge, United Kingdom) was added per 1 ml of BLEB. Incubation was subsequently continued at 30 ± 2°C and enrichments were tested for *Listeria* spp and *L. monocytogenes* after 24 h and 48 h of incubation. At each the 24 h and the 48 h time point, 50 µl of the enrichment were plated on both *L. monocytogenes* plating medium (LMPM; R&F Products, Inc., Downers Grove, IL) and modified Oxford medium (MOX; BD) agar plates, which were incubated for 48 h at 35 ± 2°C and 30 ± 2°C, respectively. After 48 h, the plates were compared to a positive control to identify putative *Listeria* spp. (gray- green colonies with black halos on MOX or white colonies on LMPM) or putative *L. monocytogenes* (turquoise colonies on LMPM agar). Colonies that appeared to be *Listeria* spp. were sub-streaked onto LMPM agar plates, followed by incubation for 48 h. For each sample, up to four turquoise colonies (i.e., presumptive *L. monocytogenes*) and two white (i.e., presumptive *Listeria* spp.) were plated onto BHI agar plates, which were incubated for 24 h at 37 ± 2°C; these isolates were frozen at -80°C in 15% glycerol.

For each sample, one presumptive *Listeria* spp. colony was characterized by sequencing of a fragment of the *sigB* gene. If a sample yielded both turquoise and white colonies on LMPM, one colony of each color was further characterized. As previously described, a single colony was used to prepare a lysate, which was subsequently used to amplify a 660 bp portion of *sigB* (Nightingale et al., 2005; den Bakker et al., 2010). PCR products were sequenced (in both directions) using Sanger sequencing performed at the Cornell University Life Sciences Core Laboratories Center. The resulting sequences were compared with an internal database (Food Safety Laboratory, Cornell University, Ithaca, NY) using BioEdit (Hall, 2001) as previously described (Nightingale et al., 2007; Ivy et al., 2012) in order to assign each isolate a *sigB* allelic type (AT); isolates with different ATs show 660 nucleotide (nt) partial *sigB* sequences that differ by at least a single nt. Using our existing database, *sigB* sequence data can be used to assign a *Listeria* species to each isolate; this

approach has been shown in a number of studies to allow for reliable speciation of *Listeria* spp. isolates (Nightingale et al., 2007; Sauders et al., 2012; Simmons et al., 2014). Two-sided Fisher's exact tests were used to determine if a *sigB* AT is significantly associated with a particular facility.

Pulsed-Field Gel Electrophoresis of *Salmonella* and *L. monocytogenes* isolates. PFGE was performed, according to the PulseNet protocol (Graves and Swaminathan, 2011; Centers for Disease Control 2013a; Centers for Disease Control, 2013b), on one *Salmonella* and one *L. monocytogenes* isolate from each sample that tested positive for these pathogens. PFGE was performed with one restriction enzyme (*XbaI*) for *Salmonella* and two restriction enzymes (*AscI* and *ApaI*) for *L. monocytogenes*. Electrophoresis was performed on CHEF DR-II and CHEF- Mapper instruments (Bio-Rad, Hercules, CA) using optimized running parameters outlined in the PulseNet protocols.

PFGE patterns were analyzed using Bionumerics (Applied Maths, Austin, TX). PFGE patterns were compared to a *Listeria* database containing patterns for approximately 3,500 isolates (representing > 500 different patterns for each restriction enzyme) or a *Salmonella* database containing patterns for about 4,000 *Salmonella* isolates (representing approximately 1,000 different *XbaI* patterns). Two-sided Fisher's exact tests were performed on *L. monocytogenes* PFGE patterns and processing facilities to determine whether a PFGE pattern was significantly associated with a particular facility.

Statistical analysis of *Listeria* spp. prevalence by zone. We analyzed the prevalences of *Listeria* spp. and *L. monocytogenes* in Zones 2, 3, and 4 by fitting, for each response, a logistic mixed-effect regression using zone as a fixed effect and facility as a random effect with the lme4 and lsmeans packages in R (R Core Team, 2013; Bates et al., 2014; Lenth, 2016).

Validation of EMPs. Following at least six months of routine sampling, validation of the EMPs was conducted at seven facilities (A, E, F, G, H, I, and J). Validation sampling was performed to determine whether the *Listeria* spp. frequency detected during routine sampling correctly estimated the *Listeria* spp. frequency in a given facility that could be detected by an independent entity. Validation sampling was performed by individuals who did not conduct the routine sampling but had experience in environmental sampling; these individuals were allowed to select any additional Zone 2, 3, or 4 sites in the facilities sampled, independent of the sites listed in the EMPs; collection and testing of sponge samples was performed as described above. Sample size calculations were performed to determine how many environmental samples needed to be collected to have an 80% power to determine whether the *Listeria* spp. detection frequency during validation sampling was no more than twice the frequency calculated based on the routine sampling. These sample size calculations used the routine sampling results to estimate the minimum detectable difference in the validation sampling using a normal approximation to the binomial distribution (Millard, 2013; R Core Team, 2013). For example, if the *Listeria* spp. frequency observed during regular sampling was 15%, we estimated the number of samples that needed to be collected to be 80% confident that the "true" frequency was < 30%. If the observed frequency was < 5%, we estimated the number of samples needed to have an 80% power to determine whether the "true" frequency was < 10%. A two-sided Fisher's exact test was performed on the routine testing and validation sampling results to determine whether the *Listeria* spp. frequency observed during routine sampling was significantly different from the *Listeria* spp. frequency observed during the validation testing.

RESULTS AND DISCUSSION

In this study, 4,613 samples (183 raw milk samples and 4,430 environmental sponge samples) were collected from nine different small cheese processing facilities, including

seven facilities classified as farmstead, and tested for *Listeria* spp. In addition, 625 samples collected from five facilities were tested for *Salmonella*; *Salmonella* testing was only conducted for a subset of facilities due to low *Salmonella* prevalence (see below for details). Overall, our data show that *Salmonella* is only rarely found in small cheese processing facility environments, while *Listeria* spp. (including *L. monocytogenes*) are more commonly found in these facilities, representing both persistent and transient strains. We also showed that the approach used here for validation of EMPs allowed for identification of two facilities that had significantly higher prevalence in the validation sampling as compared to the routine sampling. The remaining five facilities where validation sampling was performed, all showed numerically lower prevalence in the validation sampling than the routine sampling (including two facilities with statistically significantly lower prevalence in the validation sampling); this suggests that appropriately implemented EMPs may reduce the risk of pathogen detection by third parties. Importantly, the data created here provide valuable benchmarking information for both *Listeria* spp. and *Salmonella* EMPs.

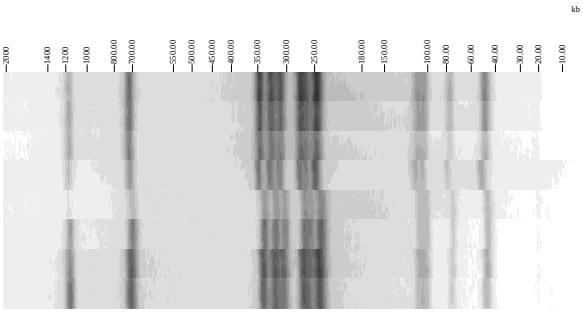
***Salmonella* was only isolated from one facility and was likely introduced from the farm environment.** Overall, 625 total samples were tested for *Salmonella*, including; (i) 588 environmental samples (from facilities A, C, D, E, and F); (ii) 30 raw milk samples (from raw milk collected at these same facilities) and; (iii) 7 follow-up samples collected at facility C. Only one environmental sample (0.2%) and three raw milk samples (10%) tested positive; all four of these samples were associated with Facility C. By comparison, Van Kessel et al. (2004) found *Salmonella* in 22/861 (2.6%) bulk tank milk samples in a study that evaluated prevalence of pathogens in bulk tank milk across the U.S. A review paper reported *Salmonella* isolation rates between 0.2% and 8.9% for bulk tank milk samples collected from dairy farms across the U.S. (Oliver et al., 2005), also consistent with the results reported here.

As *Salmonella*-positive raw milk and environmental samples were only found in

Facility C (a farmstead operation), we collected seven follow-up environmental samples from the milking parlor of the nearby farm that provided raw milk to Facility C; five of these samples tested positive for *Salmonella*. PFGE analysis of one *Salmonella* isolate from each of the nine positive samples from facility C (three raw milk, one processing facility environmental sample, five follow-up samples) yielded the same pattern with restriction enzyme *XbaI* (pattern CU-213; Figure 2.1). This pattern matched a PFGE pattern that had previously been reported to be common among *Salmonella enterica* subsp. *enterica* serotype Cerro isolates, and unique for this serotype (Mammina et al., 2000; Cummings et al., 2010; Hoelzer et al., 2011; Tewari et al., 2012; Haley et al., 2014; Rodriguez-Rivera et al., 2014). We thus concluded that the isolates collected at Facility C and the associated farm represent serotype Cerro. Importantly, *Salmonella* Cerro has emerged as one of the most common *Salmonella* serotypes associated with dairy cattle in the U.S. (Van Kessel et al., 2007; Cummings et al., 2010) and is regularly isolated on dairy farms (Fossler et al., 2004; Rodriguez et al., 2006; Van Kessel et al., 2007; Cummings et al., 2010; Hoelzer et al., 2011). For example, in one study in New York, *Salmonella* Cerro was found in 20/44 (45%) of dairy farms that tested positive for *Salmonella* (Cummings et al., 2010). Interestingly, despite the common isolation of this serotype in cattle, human salmonellosis outbreaks linked to this serotype are rare (even though one foodborne outbreak involving two Arkansas prisons occurred in 2012 [Gicquelais et al., 2014]). One recent study suggested that at least some of the *Salmonella* Cerro strains isolated from cattle show mutations that may be responsible for reduced human virulence (Rodriguez-Rivera et al., 2014).

Despite the fact that *Salmonella* Cerro may show reduced human virulence, our data suggest that environmental sources of *Salmonella* in dairy processing facilities need to be considered when assessing food safety risks. While environmental sources and cross- contamination of RTE products with *Salmonella* have been linked to salmonellosis outbreaks in other products (Sheth et al., 2011; Russo et al., 2013), outbreaks of *Salmonella* in cheese have almost always

been linked to raw milk as a contamination source (Desenclos et al., 1993; Cody et al., 1999). However, a 1989 salmonellosis outbreak from cheese was attributed to contamination from environmental sources or an infected worker in the processing environment (Hedberg et al., 1992). While we appreciate that the case reported here represents a single facility, it is tempting to speculate that the source of the *Salmonella* Cerro environmental contamination in the processing facility may have been the adjacent farm with introduction though either raw milk or fomites (e.g., shoes). This is supported by other studies that indicate a considerable prevalence of *Salmonella* on farms; for example, one study reported 17.9% *Salmonella* prevalence among 624 rectal swabs, feed, soil, and bedding samples collected from dairy farms (Rodriguez et al., 2006). Furthermore, Cummings et al. (2010) found 44/57 (77%) dairy herds in New York State had environmental or fecal samples test positive for *Salmonella* during the study period. Overall, these data suggest that dairy farms may be a source of *Salmonella* that could be introduced into the environment of farmstead processing facilities.



Key	Serovar	Date
FSL A5-0214	Cerro	7/2013
FSL A5-0218	Cerro	7/2013
FSL A5-0271	Cerro	8/2013
FSL A5-0285	Cerro	9/2013
FSL A5-0290	Cerro	9/2013
FSL A5-0295	Cerro	9/2013
FSL A5-0300	Cerro	9/2013
FSL A5-0302	Cerro	9/2013

FIGURE 2.1: *XbaI* PFGE patterns for selected *Salmonella* isolates obtained from samples associated with Facility C, including isolates from raw milk, the processing facility environment, and follow-up samples obtained from the milking parlor of the adjacent farm. All isolates yielded *XbaI* pattern CU-213, which matched with previously subtyped *Salmonella* Cerro isolates. The date indicates the day of sample collection.

***Listeria* spp. and *L. monocytogenes* prevalence in 183 raw milk samples was 12% and 8%, respectively.** Among 183 raw milk samples collected from eight facilities (A, C, D, E, F, G, H, and J), 22 and 14 samples were positive for *Listeria* spp. and *L. monocytogenes*, respectively. The positive raw milk samples came from two facilities, including Facility A (20 *Listeria* spp. and 13 *L. monocytogenes* positive samples out of 73 samples) and Facility C (2 *Listeria* spp. and 1 *L. monocytogenes* positive samples out of 19 samples). *Listeria* spp. detection did differ significantly between facilities ($P < 0.05$; Fishers exact test) with Facility A showing significantly higher *Listeria* spp. ($P = 0.0003$) as compared to the other facilities. Interestingly, Facility A was unique in that it receives milk from 15 dairy farms utilizing milking practices and farming methods with minimal mechanization. Our data thus suggest that monitoring of raw milk for *Listeria* spp. may be able to identify raw milk supplies that represent an increased risk of *L. monocytogenes* introduction.

In addition to 14 raw milk samples positive for *L. monocytogenes*, the *Listeria* species identified among the raw milk isolates included *L. innocua*, *L. seeligeri*, and *L. welshimeri* (found in six, two, and one raw milk samples respectively, see Table 2.2). The raw milk *Listeria* spp. and *L. monocytogenes* prevalence found here is consistent with previous studies (Waak et al., 2002; Van Kessel et al., 2004; Oliver et al., 2005; United States Department of Agriculture Veterinary Services, 2009; Van Kessel et al., 2011). For example, a 2004 study of 860 bulk tank milk samples collected from dairy farms in 21 states detected *Listeria* spp. and *L. monocytogenes* in 10.4% and 6.5% of samples, respectively (Van Kessel et al., 2004). Review articles have suggested isolation of *L. monocytogenes* in raw milk ranging from 1.0% to 12.6% (Oliver et al., 2005; Kousta et al., 2010).

TABLE 2.2: Occurrence of different *sigB* allelic types for *Listeria* spp. isolates from each processing facility[a]

			No. *Listeria* isolates from facility[c]													
			Environmental samples										Raw milk samples			
Species		*sigB* allelic type	A	C	D	E	F	G	H	I	J	subtotal	A	C	subtotal	Total
L. booriae		NA[b]	5*	-	-	-	-	-	-	-	-	5	-	-	0	5
	subtotal		5	0	0	0	0	0	0	0	0	5	0	0	0	5
L. innocua		6	1	1	-	-	-	-	-	-	-	2	-	-	0	2
		23	1	-	-	-	-	-	-	-	-	1	-	-	0	1
		26	-	1	-	-	-	4	17*	-	3	25	1	-	1	26
		30	2	-	-	-	-	-	-	-	-	2	2	-	2	4
		31	2	2	-	-	-	2	6	-	6*	18	2	1	3	21
		37	1	-	-	2	-	-	-	1	-	4	-	-	0	4
		45	1	-	-	-	-	-	-	-	-	1	-	-	0	1
		71	2	-	-	-	-	6*	-	-	-	8	-	-	0	8
		94	-	-	-	1	-	-	-	-	-	1	-	-	0	1
		116	-	1	-	117*	-	-	-	-	-	118	-	-	0	118
		125	1	-	-	-	-	-	-	-	-	1	-	-	0	1
		162	-	1	-	-	-	-	-	-	-	1	-	-	0	1
	subtotal		11	6	0	120	0	12	23	1	9	182	5	1	6	188
L. newyorkensis		NA[b]	1	12*	-	-	-	-	-	-	-	13	-	-	0	13
	subtotal		1	12	0	0	0	0	0	0	0	13	0	0	0	13
L. seeligeri		1	-	-	-	-	-	-	1	-	-	1	-	-	0	1
		3	-	-	-	-	-	-	-	-	-	0	1	-	1	1
		7	-	-	-	-	-	7*	-	-	-	7	-	-	0	7
		12	-	-	-	-	-	1	1	-	-	2	-	-	0	2
		20	-	-	-	-	4*	-	-	-	-	4	-	-	0	4
		72	-	-	-	-	-	-	-	-	-	0	1	-	1	1
		163	-	-	-	-	-	-	-	-	1	1	-	-	0	1
	subtotal		0	0	0	0	4	8	2	0	1	15	2	0	2	17
L. welshimeri		89	-	-	-	-	-	-	-	-	-	0	1	-	1	1
	subtotal		0	0	0	0	0	0	0	0	0	0	1	0	1	1

[a] Five isolates were not available for subtyping. [b] NA = not applicable; *L. booriae* and *L. newyorkensis* isolates were characterized using a *sigB* primer set that does not allow for assignment of sigB allelic types [c] A * indicates allelic types (or species for *L. booriae* or L. *newyorkensis*) that are significantly overrepresented in a given facility ($P < 0.05$).

***Listeria* spp. and *L. monocytogenes* prevalence in 4,430 environmental samples was 6.03% and 1.35%, respectively.** Overall, a total of 4,430 environmental sponge samples collected during processing were tested for *Listeria* spp. and *L. monocytogenes*; this number included both samples collected as part of routine sampling (n = 3,935) and as part of validation sampling (n = 495). The number of samples collected by facility ranged from 202 (Facility D, a seasonal processing facility) to 1,000 (Facility E, a large year-round processing facility). *Listeria* spp. and *L. monocytogenes,* respectively, were detected in 6.03% (267/4,430) and 1.35% (60/4,430) of environmental samples. *Listeria* spp. were detected in samples from all nine facilities; prevalence ranged from 0.2% (Facility I) to 14% (Facility H) (Table 2.1). *L. monocytogenes* was detected in samples from seven facilities; prevalence ranged from 0.7% (Facility E) to 5.8% (Facility C) (Table 2.1). Importantly, the timing of sample collection can have a considerable effect on the likelihood of *Listeria* spp. and *L. monocytogenes* detection. A recent document from the Innovation Center for U.S. Dairy recommends that samples be collected at least four hours into processing (Wilkin et al., 2015); use of this specific guideline for sample collection during processing could have increased *Listeria* detection and should be used in future studies.

Similar to our study, a number of other studies have reported presence of *L. monocytogenes* in the environments of dairy processing facilities throughout the world (Pritchard et al., 1995; Kabuki et al., 2004; Ho et al., 2007; Fox et al., 2011; Almeida et al., 2013; Leong et al., 2014); the *L. monocytogenes* prevalence in these other studies ranged from 2.7% (Ho et al., 2007) to 25.6% (Almeida et al., 2013). Additionally, other studies have shown that processing facilities located on or near dairy farms have an additional risk factor, as *L. monocytogenes* has been isolated from farm environments (Fox et al., 2009; Mohammed et al., 2009). Direct comparisons between studies are difficult to make, as *L. monocytogenes* detection will depend hugely on the specifics of the sampling plan and the sampling methodology (e.g., types of sampling sites selected; sampling devices used [sponges

versus swabs]). Importantly, the *L. monocytogenes* prevalence found here in small cheese processing facilities is considerably lower than that reported for urban environments in New York State (7.5 %) (Sauders et al., 2006) and soil samples collected on farms in the Northeastern U.S. (23.8%) (Nightingale et al., 2004).

In addition to *L. monocytogenes* (isolated from 60 samples), isolates were identified as *L. innocua*, *L. seeligeri*, *L. newyorkensis*, and *L. booriae* (182, 15, 13, and 5 samples, respectively). While *L. innocua* and *L. seeligeri* were found in samples from seven and four facilities respectively, *L. booriae* and *L. newyorkensis* were only found in one and two facilities, respectively. By comparison, a study of samples collected from urban and natural environments in the same region where the facilities sampled here were located, identified *L. seeligeri*, *L. welshimeri*, and *L. innocua*, as the most common *Listeria* spp. (isolated from 234, 74, and 50 of 1,805 samples, respectively) (Sauders et al., 2012). Interestingly, this previous study reported that *L. seeligeri* and *L. welshimeri* were significantly more common among natural environments, while *L. innocua* and *L. monocytogenes*, the two species most commonly found here, were significantly more common among urban environments. While it is tempting to speculate that this suggests an adaptation of *L. innocua* and *L. monocytogenes* to "built environments," future work will be needed to determine whether there are any species-specific phenotypic characteristics that would provide mechanistic support for this hypothesis.

Validation sampling indicates that for most facilities, routine sampling does not underestimate *Listeria* spp. prevalence. In order to assess whether the routine environmental sampling programs implemented in the facilities included in this study provided an appropriate assessment of the *Listeria* spp. presence in these facilities, we performed "validation" sampling at seven facilities (two facilities were not included as they ceased production). Validation is important to ensure that sampling plans implemented at a

given facility provide an accurate and representative assessment of target organism prevalence in that facility (Mayes, 1999; Sampers et al., 2012; Le et al., 2014) *(41, 48, 63)*. This "validation sampling" was performed by external experts (see "Materials and Methods"), who collected between 50 and 150 samples in each facility (see Table 2.3). Statistical analysis comparing the *Listeria* spp. frequency observed in the routine sampling (prior to the validation sampling) and in the validation sampling showed that validation prevalence was significantly lower than the routine prevalence for Facilities A and H ($P = 0.0457$ and $P = 0.0026$, respectively), but was significantly higher for Facilities F and J ($P = 0.0021$ and $P = 0.0015$, respectively). This suggests that the routine sampling program in Facilities F and J may not have appropriately assessed *Listeria* spp. presence in these facilities, which could be due to a variety of factors including exclusion of important sampling sites or issues with sampling technique. Alternatively, validation sampling could have, by chance, occurred at a time of an incident that led to an increased *Listeria* spp. prevalence. Interestingly, both facilities where validation sampling found a significantly higher prevalence than routine sampling were among the three facilities that had a very low *Listeria* prevalence (< 1%) during routine sampling (Table 2.3). Importantly, we did not consistently or formally observe the routine swabbing procedures in the facilities studied and thus do not have scientifically valid data that would provide insight as to why routine swabbing in some facilities yielded lower detection rates than validation swabbing. Future studies addressing this issue may be valuable.

This part of our study illustrates an approach that could be broadly used to validate whether environmental monitoring programs correctly assess *Listeria* spp. presence in a given facility. The application of a statistically-derived validation program was similar to a previous report that applied an approach using binomial probabilities to validate several data sets involving detection of chemical and biological hazards in product samples (Lee et al., 2016). Lee et al. (2016) found that overall, and in most individual samplings, violation rates in the

validation data were lower than the violation rates observed in the previously collected data, which is similar to our data that showed, for most facilities, lower prevalence at the time of validation (as compared to the routine sampling).

TABLE 2.3: Routine and validation prevalence of *Listeria* spp. by facility

Facility	Routine prevalence in % [a]	Samples needed for validation	Samples taken for validation	Validation sampling prevalence in % [b]
A	5.1% (34/664)	150	150	1.3%* (2/150)
E	11.1% (88/795)	60	60	10.0% (6/60)
F	<0.3% (0/334)	15	50	6.0%* (3/50)
G	9.1% (19/209)	85	85	2.4% (2/85)
H	22.6 % (24/106)	25	50	4.0%* (2/50)
I	0.4% (1/222)	15	50	<2.0% (0/50)
J	0.9% (1/106)	25	50	14.0%* (7/50)

[a] Routine prevalence is based on the data that were collected prior to validation sampling; this prevalence value was used to calculate the sample sizes for the validation sampling facilities.

[b] A * indicates that prevalence of *Listeria* spp. in routine and validation samplings were significantly different ($P < 0.05$)

Samples from Zone 2 were less likely to test positive for *Listeria* spp. than samples from Zone 3 than samples from Zone 4.

During this study, we collected 1,063 Zone 2 samples, 3,009 Zone 3 samples, and 358 Zone 4 samples that were tested for *Listeria* spp. and *L. monocytogenes*. A generalized linear mixed model was fit to the prevalence data to evaluate the effect of zone on the probability that the sample will be positive for (i) total *Listeria* spp. including *L. monocytogenes* and (ii) *L. monocytogenes* specifically. Facility was a random effect while the presence or absence of any *Listeria* spp. was a fixed effect. Using lsmeans in R (Lenth, 2016), we determined the probabilities of finding *Listeria* spp. in a facility are: 2.76% (Zone 2), 3.52% (Zone 3), and 5.72% (Zone 4); probabilities of detecting *Listeria* spp. differed significantly between zones 2 and 4 ($P = 0.0046$). The probabilities of finding *L. monocytogenes* were as follows: 0.126% (Zone 2), 0.996% (Zone 3), and 3.264% (Zone 4); probabilities of detecting *L. monocytogenes* differed significantly between all zones ($P \leq 0.0100$). Overall, these results were expected as Zone 4, by definition, is outside of the processing area, and hence includes areas (e.g., offices) that are cleaned and sanitized less frequently than Zone 2 and 3 areas.

This study is one of the first that provides baseline data for *Listeria* spp. and *L. monocytogenes* prevalence by zone. This is important as a number of guidance documents (Grocery Manufacturers Association, 2014; Wilkin et al., 2015) suggest that environmental samples be classified into 4 zones, as was done here. One previous study in smoked seafood plants categorized samples as drain, non-food contact surface, employee contact surfaces, and food contact surfaces (Lappi et al., 2004). That previous study (Lappi et al., 2004) found higher *Listeria* spp. prevalence in drains and on non-food contact surfaces (likely representing Zones 2 and 3) than on food contact surfaces (Zone 1). Wulff et al. (2006) sampled Zones 1, 2, and 3 in Danish fish slaughter- and smokehouses for *L. monocytogenes* both during production and after cleaning. For the four slaughterhouses, 50% of Zone 3 samples tested positive for *L. monocytogenes* during production, while 43% and 42% of samples from Zones

2 and 1, respectively, tested positive. In the smokehouses, 41% of Zone 3 samples tested positive for *L. monocytogenes* while 24% and 15% of Zones 2 and 1, respectively, tested positive (Wulff et al., 2006). The data from the previous study (Wulff et al., 2006) shows that the prevalence of *L. monocytogenes* was lower on food contact surfaces than non-food contact surfaces; similarly, our data shows that the prevalence of *Listeria* spp. and *L. monocytogenes* gradually decreased from Zone 4 to 2.

Subtype data suggest *Listeria* spp. contamination in the small cheese processing facilities evaluated largely represents persistent strains. Characterization of all *Listeria* spp. isolates by sequencing of a 660 nt portion of the *sigB* gene allowed us to classify the 267 isolates obtained into 26 different *sigB* ATs, including 6 for *L. monocytogenes*, 12 for *L. innocua*, 7 for *L. seeligeri*, and 1 for *L. welshimeri* (Table 2.2). Isolates identified as *L. booriae* and *L. newyorkensis* could not be assigned to *sigB* ATs as a different set of primers that yields shorter sequences had to be used for these isolates (Weller et al., 2015). Overall, 13 *sigB* ATs were isolated from more than one processing facility, while the other 13 *sigB* ATs were only found in one facility (e.g., all 7 *L. seeligeri* AT 7 isolates were obtained from Facility G). To more formally determine whether certain ATs were significantly associated with different processing facilities, prevalence rates of *Listeria* spp. *sigB* ATs were compared between facilities using two-sided Fisher's exact tests (with *P*-values adjusted for multiple comparisons); for these analyses, *L. booriae* and *L. newyorkensis* were each treated as a single subtype. Among the nine facilities, four (E, F, H, and J) showed evidence for significant overrepresentation of one AT and one facility (G) showed evidence for significant overrepresentation of two ATs. For example, AT 116, which was isolated 117 times in facility E, was significantly ($P < 0.0001$) overrepresented in this facility (Table 2.2). Facility G showed evidence for overrepresentation of *L. innocua* AT 71 ($P = 0.0014$) and *L. seeligeri* AT 7 ($P < 0.0001$). In addition, *L. booriae* was significantly overrepresented in Facility A ($P =$

0.0003) and *L. newyorkensis* was significantly overrepresented in Facility C ($P < 0.0001$). Overall, among the 215 *Listeria* spp. (excluding *L. monocytogenes*) isolates characterized from environmental samples, 174 represented isolates that were overrepresented in a given facility. This suggests that the majority of *Listeria* spp. isolated in these facilities represent persistent strains, indicating that management of *Listeria* spp. persistence could considerably reduce overall *Listeria* spp. presence in these facilities.

Among the ATs that were significantly overrepresented in a given facility, all but one were found during at least three separate samplings and from more than one sampling site, further supporting that these ATs may represent *Listeria* spp. strains that are "resident" or "persistent" in these facilities. *L. seeligeri* AT 20 was the only overrepresented AT isolated on only two sampling dates (in Facility F); this AT was isolated from four sites in this facility. The most frequently isolated AT was *L. innocua* AT 116 (isolated from 117 samples in Facility E), this AT was isolated in each of the monthly samplings performed between December 2013 and March 2015, strongly supporting persistence of this strain. Sampling sites in Facility H tested positive for *L. innocua* AT 26 during eight samplings between May 2014 and March 2015. *L. seeligeri* AT 20 was isolated during two routine samplings of Facility F (August 2013 and October 2014), as well as during the validation sampling (October 2014), where three *L. seeligeri* AT 20 isolates were found. In Facility G, which had two different putatively "persistent" *Listeria* spp., *L. seeligeri* AT 7 was isolated every month between May and August 2014 and *L. innocua* AT 71 was isolated every month between October and December 2014. *L. booriae* was found in Facility A during two different samplings between August and November 2013 and *L. newyorkensis* was found in Facility C during four different samplings between July and October of 2013.

PFGE data suggest that *L. monocytogenes* contamination includes sporadic as well as persistent strains. All *L. monocytogenes* isolates were also differentiated using PFGE to

further assess persistence of this pathogen with a highly discriminatory subtyping method. Using a combination of *ApaI* and *AscI* PFGE patterns, 25 different pattern combinations were identified (Table 2.4). Among the 57 environmental isolates available for subtyping, 22 different PFGE patterns were identified, while among the 12 raw milk isolates available for subtyping, 5 different PFGE types were identified; three of these PFGE types were only found among raw milk isolates. One pattern (CU-172,425) was found among both raw milk (collected in April 2013) and environmental isolates (collected in October and December 2013) from Facility C. One other pattern (CU-258,67) was represented by 5 raw milk isolates (collected at Facility A) as well as environmental isolates from Facilities G and H (Table 2.4), suggesting that this pattern represents widely distributed PFGE pattern, at least in the environments and geographical areas studied here.

Among the 22 patterns observed among the 57 environmental isolates, 14 were found only once (8 patterns) or twice (6 patterns). Three patterns were significantly associated with a particular processing facility, including (i) pattern CU-258,67 (Facility G; $P = 0.041$), (ii) pattern CU-29,361 (Facility H; $P = 0.0032$), and (iii) pattern CU-42,551 (Facility J; $P = 0.0042$). Pattern CU-258,67 was found six times and over five different sampling times (between June and December 2014) in Facility G, supporting persistence of this strain. Pattern CU-29,361 was obtained during two sampling visits in Facility H (May and June 2014; five isolates and one isolate, respectively). While some authors proposed (Kabuki et al., 2004; Simmons et al., 2014) that only isolation over three different sampling times suggests persistence of a given strain, isolation over two sampling times may still support persistence, particularly since pattern CU-29,361 was only found in Facility H. On the other hand, all three isolates with pattern CU-42,551 in Facility J were obtained at the same sample collection date, suggesting sporadic contamination of multiple sites.

In addition to isolation of *L. monocytogenes* with patterns CU-258,67 and CU-29,361 over multiple sampling visits to a given facility, six other patterns were also found during

multiple visits to a given facility. Specifically, (i) pattern CU-54,570 was found over two samplings (May and June 2013) in Facility A, (ii) pattern CU-311,37 was found over three samplings (October, November, and December 2013) in Facility A, (iii) pattern CU-172,425 was found over two samplings (October and December 2013) in Facility C, (iv) pattern CU-11,534 was found over two samplings (July and November 2014) in Facility G, (v) pattern CU-133,39 was found over four samplings (June, July, and December 2014 and March 2015) in Facility H and (vi) pattern CU-490,155 was found over two samplings (May and July 2014) in Facility H. If we use a lenient definition of re-isolation of *L. monocytogenes* with a given PFGE pattern over two or more different sampling dates as evidence for persistence, then 31 of the 57 environmental *L. monocytogenes* isolates may represent persistent strains. With this same definition, we identified persistent *L. monocytogenes* in four of the nine facilities (A, C, G, and H). By comparison, a study in Ireland reported persistence of a given *L. monocytogenes* PFGE type in 11 of 48 processing facilities representing various food sectors (Leong et al., 2014); in that study, persistence was also defined as isolation of *L. monocytogenes* with a given PFGE pattern at least twice (over a period of six months) in the same processing facility or the associated final product. Most other studies on multiple food processing facilities also found that many if not most facilities showed evidence for *L. monocytogenes* persistence. For example, Lappi et al. (2004) found evidence for *L. monocytogenes* persistence in three of four smoked seafood facilities studied, and Simmons et al. (2014) found evidence for *L. monocytogenes* persistence in 12 of 30 retail delis sampled.

TABLE 2.4: Prevalence of *L. monocytogenes* isolates by PFGE pattern

	No. *L. monocytogenes* isolates from facility											
	Environmental samples[a]								Raw milk samples			
PFGE Pattern	A	C	D	E	G	H	J	subtotal	A	C	subtotal	Total
CU-258,67	-	-	-	-	6*	1	-	7	5	-	5	12
CU-172,425	-	6	-	-	-	-	-	6	-	1	1	7
CU-29,361	-	-	-	-	-	6*	-	6	-	-	0	6
CU-311,37	3	1	1	-	-	-	-	5	-	-	0	5
CU-133,39	-	-	-	-	-	4	-	4	-	-	0	4
CU-42,551	-	-	-	-	-	-	3*	3	-	-	0	3
CU-54,570	2	-	-	1	-	-	-	3	-	-	0	3
CU-254,53	-	-	-	-	-	-	-	0	3	-	3	3
CU-490,155	1	-	-	-	-	2	-	3	-	-	0	3
CU-11,534	-	-	-	-	2	-	-	2	-	-	0	2
CU-68,151	1	1	-	-	-	-	-	2	-	-	0	2
CU-140,36	-	2	-	-	-	-	-	2	-	-	0	2
CU-140,518	-	-	-	-	-	-	-	0	2	-	2	2
CU-181,175	-	-	-	2	-	-	-	2	-	-	0	2
CU-225,73	-	2	-	-	-	-	-	2	-	-	0	2
CU-254,446	-	2	-	-	-	-	-	2	-	-	0	2
CU-169,114	-	1	-	-	-	-	-	1	-	-	0	1
CU-225,535	1	-	-	-	-	-	-	1	-	-	0	1
CU-259,67	1	-	-	-	-	-	-	1	-	-	0	1
CU-362,10	-	-	-	1	-	-	-	1	-	-	0	1
CU-422,507	-	-	-	1	-	-	-	1	-	-	0	1
CU-425,506	1	-	-	-	-	-	-	1	-	-	0	1
CU-426,508	1	-	-	-	-	-	-	1	-	-	0	1
CU-460,533	-	-	-	-	-	-	-	0	1	-	1	1
CU-491,423	1	-	-	-	-	-	-	1	-	-	0	1
Total	12	15	1	5	8	13	3	57	11	1	12	69[b]

[a] A* indicates a significant overrepresentation of a given PFGE type in a given facility ($p<0.05$).

[b] While 73 samples tested positive for *L. monocytogenes* only 69 were available for PFGE

Our data support that *Listeria* spp. or *L. monocytogenes* persistence is common among small cheese processing facilities. Only two of the nine facilities studied here (Facilities D and I) did not show evidence for *Listeria* spp. or *L. monocytogenes* persistence. Improved control and elimination of persistent subtypes thus would likely considerably reduce *Listeria* spp. and *L. monocytogenes* contamination of processing facilities. Achieving this will require implementation of stringent environmental sampling plans with appropriate follow-up actions that can eliminate persistent strains, a strategy also known as Seek and Destroy (Malley et al., 2015). Importantly, our study also provides a road map for a validation sampling strategy that we have shown can identify facilities where routine sampling may not correctly assess *Listeria* spp. or *L. monocytogenes* contamination. While we did not collect finished product samples, future studies including both environmental and finished product samples could provide further insights in the transmission and ecology of *Listeria* in small cheese processing facilities and specifically help characterize the likelihood of cross-contamination of finished products from different environmental sources.

ACKNOWLEDGEMENTS: This project was supported by an unrestricted gift from Wegmans Foods and by the New York State Dairy Promotion Advisory Board. The authors gratefully acknowledge Maureen Gunderson for her assistance in media preparation. The authors also acknowledge the students and staff of the Food Safety Lab and Milk Quality Improvement Program, especially Aljoša Trmčić, Veronica Guariglia, Anna Sophia Harrand, Clinton Hervert, Alexander Alles, and Alphina Ho for their assistance during this project.

REFERENCES

Acciari V. A., L. Iannetti, A. Gattuso, M. Sonnessa, G. Scavia, C. Montagna, N. Addante, M. Torresi, L. Zocchi, S. Scattolini, P. Centorame, C. Marfoglia, V. A. Prencipe, and M. V. Gianfranceschi. 2015. Tracing sources of *Listeria* contamination in traditional Italian cheese associated with a US outbreak: investigations in Italy. *Epidemiol. Infect*. doi: 10.1017/S095026881500254X.

Almeida, G., R. Magalhães, L. Carneiro, I. Santos, J. Silva, V. Ferreira, T. Hogg, and P. Teixeira. 2013. Foci of contamination of *Listeria monocytogenes* in different cheese processing facilities. *Int. J. Food Microbiol.* 167:303-309.

American Cheese Society. Cheese Glossary- What is Farmstead Cheese? Available at: http://www.cheesesociety.org/i-heart-cheese/cheese-glossary/. Accessed 10 February 2016.

Andrews, W. H., A. Jacobson, and T. Hammack. 2014. FDA Bacterial Analytical Manual Chapter 5- *Salmonella*. Available at: http://www.fda.gov/Food/FoodScienceResearch/LaboratoryMethods/ucm070149.htm. Accessed 8 February 2016.

Barancelli, G. V., T. M. Camargo, N. G. Gagliardi, E. Porto, R. A. Souza, F. Campioni, J. P. Falcão, E. Hofer, A. G. Cruz, and C. A. F. Oliveira. 2014. Pulsed-field gel electrophoresis characterization of *Listeria monocytogenes* isolates from cheese manufacturing facilities in

São Paulo, Brazil. *Int. J. Food Microbiol*. 173:21-29.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2014. Lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-6. Available at: http://CRAN.R-project.org/package=lme4. Accessed 5 February 2016.

Bille, J., D. S. Blanc, H. Schmid, K. Boubaker, A. Baumgartner, H. H. Siegrist, M. L. Tritten, R. Lienhard, D. Berner, R. Anderau, M. Treboux, J. M. Ducommun, R. Malinverni, D. Genné, P. Erard, U. Waespi. 2006. Outbreak of human listeriosis associated with Tomme cheese in Northwest Switzerland, 2005. *Euro Surveill*. 11:91-93.

Carpentier, B., and O. Cerf. 2011. Review- Persistence of *Listeria monocytogenes* in food industry equipment and premises. *Int. J. Food Microbiol*. 145:1-8.

Carrasco, E., A. Morales-Rueda, and R. M. García-Gimeno. 2012. Cross-contamination and recontamination by *Salmonella* in foods: A review. *Food Res. Int.* 45:545–556.

Centers for Disease Control and Prevention (CDC). 2012. Multistate outbreak of listeriosis linked to imported Frescolina Marte brand Ricotta Salata cheese (final update). Available at: http://www.cdc.gov/listeria/outbreaks/cheese-09- 12/index.html?s_cid=fb1807. Accessed 25 November 2015.

Centers for Disease Control and Prevention (CDC). 2013. Standard Operating Procedure for PulseNet PFGE of *Listeria monocytogenes*. Available at: http://www.cdc.gov/pulsenet/PDF/listeria-pfge-protocol-508c.pdf. Accessed 23 November 2015.

Centers for Disease Control and Prevention (CDC). 2013. Standard Operating Procedure for PulseNet PFGE of *Escherichia coli* O157:H7, *Escherichia coli* non-O157:H7 (STEC), *Salmonella* serotypes, *Shigella sonneri* and *Shigella flexneri*. Available at: http://www.cdc.gov/pulsenet/PDF/ecoli-shigella-salmonella-pfge-protocol-508c.pdf. Accessed 23 November 2015.

Cody, S. H., S. L. Abbott, A. A. Marfin, B. Schulz, P. Wagner, K. Robbins, J. C. Mohle- Boetani,

and D. J. Vugia. 1999. Two outbreaks of multidrug-resistant *Salmonella* serotype Typhimurium DT104 infections linked to raw-milk cheese in Northern California. *J. Am. Med. Assoc*. 281:1805-1810.

Cummings, K. J., L. D. Warnick, M. Elton, L. D. Rodriguez-Rivera, J. D. Siler, E. M. Wright, Y. T. Grohn, and M. Wiedmann. 2010. *Salmonella enterica* serotype Cerro among dairy cattle in New York: an emerging pathogen? *Foodborne Pathog. Dis.* 7:659- 665.

den Bakker, H. C., B. N. Bundrant, E. D. Fortes, R. H. Orsi, and M. Wiedmann. 2010. A population genetics-based and phylogenetic approach to understanding the evolution of virulence in the genus *Listeria*. *Appl. Environ. Microbiol*. 76:6085-6100.

Desenclos, J.-C., P. Bouvet, E. Benz-Lemoine, F. Grimont, H. Desqueyroux, I. Rebiere, and P. A. Grimont. 1996. Large outbreak of *Salmonella enterica* serotype Paratyphi B infection caused by a goats' milk cheese, France, 1993: a case finding and epidemiological study. *BMJ* 312:91-94.

Ferreira, V., M. Wiedmann, P. Teixeira, and M. J. Stasiewicz. 2014. *Listeria monocytogenes* persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. *J. Food Prot*. 77:150-170.

Food and Drug Administration (FDA). 2013. FDA investigation summary- Multi-state outbreak of *Listeria monocytogenes* linked to certain Crave Brothers Farmstead Classic Cheeses. Available at: http://www.fda.gov/Food/RecallsOutbreaksEmergencies/Outbreaks/ucm359588.htm. Accessed 25 November 2015.

Food and Drug Administration (FDA). 2015. FSMA final rule for preventative controls for human food. Available at: http://www.fda.gov/Food/GuidanceRegulation/FSMA/ucm334115.htm. Accessed 24 November 2015.

Fossler, C. P., S. J. Wells, J. B. Kaneene, P. L. Ruegg, L. D. Warnick, J. B. Bender, S. M. Godden,

L. W. Halbert, A. M. Camppbell, and A. M. Zwald. 2004. Prevalence of *Salmonella* spp. on conventional and organic dairy farms. *J. Am. Vet. Med. Assoc*. 225:567-573.

Fox, E., K. Hunt, M. O'Brien, and K. Jordan. 2011. *Listeria monocytogenes* in Irish Farmhouse cheese processing environments. *Int. J. Food Microbiol*. 145:S39-S45.

Fox, E., T. O'Mahony, M. Clancy, R. Dempsey, M. O'Brien, and K. Jordan. 2009. *Listeria monocytogenes* in the Irish dairy farm environment. *J. Food Prot*. 72:1450-1456.

Gaulin C., D. Ramsay, and S. Bekal. 2012. Widespread listeriosis outbreak attributable to pasteurized cheese, which led to extensive cross-contamination affecting cheese retailers,

Quebec, Canada, 2008. *J. Food Prot.* 75:71-78.

Gicquelais, R. E., J. F. Morris, H. S. Matthews, L. Gladden, H. Safi, C. Grayson, R. B. Slayton, A. E. Newton, R. Bordonaro, J. G. Wheeler, N. Smith, S. A. Bosch, and D. T. Haselow. 2014. Multiple-serotype *Salmonella* outbreaks in two state prisons- Arkansas 2012. *MMWR Morb. Mortal. Wkly*. Rep. 63:169-173.

Gould L. H., E. Mungai, and C. B. Behravesh. 2014. Outbreaks attributed to cheese: Differences between outbreaks caused by unpasteurized and pasteurized dairy products. *Foodborne Pathog Dis*. 11:545-551.

Graves, L. M., and B. Swaminathan. 2001. PulseNet standardized protocol for subtyping *Listeria monocytogenes* by macrorestriction and pulsed-field gel electrophoresis. *Int. J. Food Microbiol*. 65:55-62.

Grocery Manufacturers Association. 2014. *Listeria monocytogenes* guidance on environmental monitoring and corrective actions in at-risk foods. Available at: http://ucfoodsafety.ucdavis.edu/files/208833.pdf. Accessed 8 February 2016.

Haley, B. J., Y. Luo, C. Wang, J. Pettengill, M. Allard, E. Brown, J. S. Karns, and J. A. Van Kessel. 2014. Genome sequences of eight *Salmonella enterica* subsp. *enterica* serovars isolated from a single dairy farm. *Genome Announc*. doi: 10.1128/genomeA.00082-14.

Hall, T. A.. 2001. BioEdit: an important software for molecular biology. *GERF Bull. Biosci*. 2:60-61.

Hedberg, C. W., J. A. Korlath, J.-Y. D'Aoust, K. E. White, W. L. Schell, M. R. Miller, D. N. Cameron, K. L. MacDonald, and M. T. Osterholm. 1992. A multistate outbreak of *Salmonella* Javiana and *Salmonella* Oranienburg infections due to consumption of contaminated cheese. *J. Am. Med. Assoc*. 268:3203-3207.

Heiman, K. E., V. B. Garalde, M. Gronostaj, K. A. Jackson, S. Beam, L. Joseph, A. Saupe, E. Ricotta, H. Waechter, A. Wellman, M. Adams-Cameron, G. Ray, A. Fields, Y. Chen, A. Datta, L. Burall, A. Sabol, Z. Zucerova, E. Trees, M. Metz, P. LeBlanc, S. Lance, P. M. Griffin, R. V. Tauxe, and B. J. Silk. 2015. Multistate outbreak of listeriosis caused by imported cheese and evidence of cross-contamination of other cheeses, USA, 2012. *Epidemiol. Infect*. doi: 10.1017/S095026881500117X.

Hennessy, T. W., C. W. Hedberg, L. Slutsker, K. E. White, J. M. Besser-Wiek, M. E. Moen, J. Feldman, W. W. Coleman, L. M. Edmonson, K. L. MacDonald, and M. T. Osterholm. 1996. A national outbreak of *Salmonella enteritidis* infections from ice cream. *N. Engl. J. Med*. 334:1281-1286.

Hitchins, A. D., K. Jinneman, and Y. Chen. January 2016. FDA Bacterial Analytical Manual Chapter 10- Detection and enumeration of *Listeria monocytogenes*. Available at: http://www.fda.gov/Food/FoodScienceResearch/LaboratoryMethods/ucm071400.htm. Accessed 8 February 2016.

Ho, A. J., V. R. Lappi, and M. Wiedmann. 2007. Longitudinal monitoring of *Listeria monocytogenes* contamination patterns in a farmstead dairy processing facility. *J. Dairy Sci*. 90:2517-2524.

Hoelzer, K., K. J. Cummings, E. M. Wright, L. D. Rodriguez-Rivera, S.E. Roof, A. I. Moreno Switt, N. Dumas, T. Root, D. J. Schoonmaker-Bopp, Y. T. Grohn, J. D. Siler, L. D. Warnick, D. D. Hancock, M. A. Davis, and M. Wiedmann. 2011. *Salmonella* Cerro

isolated over the past twenty years from various sources in the US represent a single predominant pulsed-field gel electrophoresis type. *Vet. Microbiol.* 150:389-393.

Ivy, R. A., M. L. Ranieri, N. H. Martin, H. C. den Bakker, B. M. Xavier, M. Wiedmann, and K. J. Boor. 2012. Identification and characterization of psychrotolerant sporeformers associated with fluid milk production and processing. *Appl. Environ. Micriobiol*. 78:1853-1864.

Kabuki, D. Y., A. Y. Kuaye, M. Wiedmann, and K. J. Boor. 2004. Molecular subtyping and tracking of *Listeria monocytogenes* in Latin-style fresh-cheese processing plants. *J. Dairy Sci*. 87:2803-2812.

Kim, J. S., G. G. Lee, J. S. Park, Y. H. Jung, H. S. Kwak, S. B. Kim, Y. S. Nam, and S. Kwon. 2007. A novel multiplex PCR assay for rapid and simultaneous detection of five pathogenic bacteria: *Escherichia coli* O157:H7, *Salmonella*, *Staphylococcus aureus*, *Listeria monocytogenes*, and *Vibrio parahaemolyticus*. *J. Food Prot*. 70:1656-1662.

Kousta, M., M. Mataragas, P. Skandamis, and E. H. Drosinos. 2010. Prevalence and sources of cheese contamination with pathogens at farm and processing levels. *Food Control* 21:805–815.

Lappi, V. R., J. Thimothe, K. K. Nightingale, K. Gall, V. N. Scott, and M. Wiedmann. 2004. Longitudinal studies on *Listeria* in smoked fish plants: Impact of intervention strategies on contamination patterns. *J. Food Prot.* 67:2500-2514.

Le, S., W. Bazger, A. R. Hill, and A. Wilcock. 2014. Awareness and perceptions of food safety of artisan cheese makers in Southwestern Ontario: A qualitative study. *Food Control* 41:158-167.

Lee, K-M., T. J. Herrman, and S. Y. Dai. 2016. Application and validation of a statistically derived risk-based sampling plan to improve efficiency of inspection and enforcement. *Food Control* 64:135-141.

Lenth, R. V. 2016. Least-squares means: the R package lsmeans. *J. Stat. Softw*. 69:1-33.

Leong, D., A. Alvarez-Ordóñez, and K. Jordan. 2014. Monitoring occurance and persistence of

Listeria monocytogenes in foods and food processing environments in the Republic of Ireland. *Front. Microbiol*. 5:436. doi: 10.3389/fmicb.2014.00436.

Magalhães, R., G. Almeida, V. Ferreira, I. Santos, J. Silva, M. M. Mendes, J. Pita, G. Mariano, I.

Mâncio, M. M. Sousa, J. Farber, F. Pagotto, and P. Teixeira. 2015. Cheese- related listeriosis outbreak, Portugal, March 2009 to February 2012. *Euro Surveill*. 20: pii=21104.

Malley, T. J. V., J. Butts, and M. Wiedmann. 2015. Seek and destroy process: *Listeria monocytogenes* process controls in the ready-to-eat meat and poultry industry. *J. Food Prot*. 78:436-445.

Mammina, C., L. Cannova, S. Carfi Pavia, and A. Nastasi. 2000. Endemic presence of *Salmonella enterica* serotype *Cerro* in southern Italy. *Euro Surveill*. 5:84-86.

Mayes, T. 1999. How can the principles of validation and verification be applied to hazard analysis? *Food Control*. 10:277-279.

Melo, J., P. W. Andrew, and M. L. Faleiro. 2015. *Listeria monocytogenes* in cheese and the dairy environment remains a food safety challenge: The role of stress responses. *Food Res. Int*. 67:75-90.

Miettinen, M. K., K. J. Bjorkroth, and H. J. Korkeala. 1999. Characterization of *Listeria monocytogenes* from an ice cream plant by serotyping and pulsed-field gel electrophoresis. *Int. J. Food Microbiol*. 46:187-192.

Millard S. P.. 2013. *EnvStats: An R Package for Environmental Statistics.* Springer- Verlag, New York. ISBN 978-1-4614-8455-4.

Mohammed, H. O., K. Stipetic, P. L. McDonough, R. N. Gonzalez, D. V. Nyadam, and E. R. Atwill. 2009. Identification of potential on-farm sources of *Listeria monocytogenes* in herds of dairy cattle. *Am. J. Vet. Res.* 70:383-388.

Nightingale, K. K., L. Bovell, A. Grajczyk, and M. Wiedmann. 2007. Combined *sigB* allelic typing and multiplex PCR provide improved discriminatory power and reliability for

Listeria monocytogenes molecular serotyping. *J. Microbiol. Methods*. 68:52-59.

Nightingale, K. K., Y. H. Schukken, C. R. Nightingale, E. D. Fortes, A. J. Ho, Z. Her, Y. T. Grohn, P. L. McDonough, and M. Wiedmann. 2004. Ecology and transmission of *Listeria monocytogenes* infecting ruminants and in the farm environment. *Appl. Environ. Microbiol*. 70:4458-4467.

Nightingale, K.K., K. Windham, and M. Wiedmann. 2005. Evolution and molecular phylogeny of *Listeria monocytogenes* isolated from human and animal listeriosis cases and foods. *J. Bacteriol*. 187:5537-5551.

Oliver, S. P., B. M. Jayarao, and R. A. Almeida. 2005. Foodborne pathogens in milk and the dairy farm environment: food safety and public health implications. *Foodborne Pathog. Dis*. 2:115-129.

Parisi, A., L. Latorre, R. Fraccalvieri, A. Miccolupo, G. Normanno, M. Caruso, and G. Santagada. 2013. Occurrence of *Listeria* spp. in dairy facilities in Southern Italy and molecular subtyping of isolates using AFLP. *Food Control* 29:91-97.

Pritchard, T. J., K. J. Flanders, and C. W. Donnelly. 1995. Comparison of the incidence of *Listeria* on equipment versus environmental sites within dairy processing plants. *Int. J. Food Microbiol*. 26:375-384.

R Core Team. 2013. R: a language and environment for statistical computing. 10.1. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0.

Rodriguez, A., P. Pangloli, H. A. Richards, J. R. Mount, and F. A. Draughon. 2006. Prevalence of *Salmonella* in diverse environmental farm samples. *J. Food Prot*. 69:2576- 2580.

Rodriguez-Rivera, L. D., A. I. Moreno Switt, L. Degoricija, R. Fang, C. A. Cummings, M. R. Furtado, M. Wiedmann, and H. C. den Bakker. 2014. Genomic characterization of *Salmonella* Cerro ST367, an emerging *Salmonella* subtype in cattle in the United States. *BMC Genomics* 15:427-437.

Russo, E. T., G. Biggerstaff, R. M., Hoekstra, S. Meyer, N. Patel, B. Miller, and R. Quick. 2013.

A recurrent, multistate outbreak of *Salmonella* serotype Agona infections associated with dry, unsweetened cereal consumption, United States, 2008. *J. Food Prot*. 76:227-230.

Sampers, I., H. Toyofuki, P. A. Luning, M. Uyttendaele, and L. Jacxsens. 2012. Semi- quanitative study to evaluate the performance of a HACCP-based food safety management system in Japanese milk processing plants. *Food Control* 23:227-233.

Sauders, B. D., M. Z. Durak, E. Fortes, K. Windham, Y. Schukken, A. J. Lembo, B. Akey, K. K. Nightingale, and M. Wiedmann. 2006. Molecular characterization of *Listeria monocytogenes* from natural and urban environments. *J. Food Prot.* 69:93-105.

Sauders, B. D., J. Overdevest, E. Fortes, K. Windham, Y. Schukken, A. Lembo, and M. Wiedmann. 2012. Diversity of *Listeria* species in urban and natural environments. *Appl. Environ. Mirobiol*. 78:4420-4433.

Sheth, A. N., M. Hoekstra, N. Patel, G. Ewald, C. Lord, C. Clarke, E. Villamil, K. Niksich, C. Bopp, T. Nguyen, D. Zink, and M. Lynch. 2011. A national outbreak of *Salmonella* serotype Tennessee infections from contaminated peanut butter: a new vehicle for salmonellosis in the United States. *Clin. Infect. Dis*. 53:356-362.

Simmons, C., M. J. Stasiewicz, E. Wright, S. Warchocki, S. Roof, J. R. Kause, N. Bauer, S. Ibrahim, M. Wiedmann, and H. F. Oliver. 2014. *Listeria monocytogenes* and *Listeria* spp. contamination patterns in retail delicatessen establishments in three U. S. states. *J. Food Prot*. 77:1929-1939.

Stessl, B., M. Fricker, E. Fox, R. Karpiskova, K. Demnerova, K. Jordan, M. Ehling- Schulz, and M. Wagner. 2014. Collaborative survey on the colonization of different types of cheese-processing facilities with *Listeria monocytogenes*. *Foodborne Pathog. Dis.* 11:8-14.

Tewari, D., C. H. Sandt, D. M. Miller, B. M. Jayarao, and N. M. M'ikanatha. 2012. Prevalence of *Salmonella* Cerro in laboratory-based submissions of cattle and comparison with human infections in Pennsylvania, 2005-2010. *Foodborne Pathog. Dis.* 9:928-933.

Tompkin, R. B. 2002. Control of *Listeria monocytogenes* in the food-processing environment. *J.*

Food Prot. 65:709-725.

Unnerstad, H., E. Bannerman, J. Bille, M. Danielsson-Tham, W. Tham, and E. Waak. 1996. Prolonged contamination of a dairy with *Listeria monocytogenes*. *Neth. Milk Dairy J.* 50:493-499.

United States Department of Agriculture (USDA) Veterinary Services. 2009. Prevalence of *Salmonella* and *Listeria* in bulk tank milk and in-line filters on U.S. Dairies, 2007. Available at: https://www.aphis.usda.gov/animal_health/nahms/dairy/downloads/dairy07/Dairy07_is_SalList.pdf. Accessed 10 February 2016.

Van Kessel, J. S., J. S. Karns, L. Gorski, B. J. McCluskey, and M. L. Perdue. 2004. Prevalence of Salmonellae, *Listeria monocytogenes*, and fecal coliforms in bulk tank milk on US dairies. *J. Dairy Sci*. 87:2822-2830.

Van Kessel, J. S., J. S. Karns, J. E. Lombard, and C. A. Koprai. 2011. Prevalence of *Salmonella enterica*, *Listeria monocytogenes*, and *Escherichia coli* virulence factors in bulk tank milk and in-line filters from U.S. dairies. *J. Food. Prot*. 74:759-768.

Van Kessel, J. S., J. S. Karns, D. R. Wolfgang, E. Hovingh, and Y. H. Schukken. 2007. Longitudinal study of a clonal, subclinical outbreak of *Salmonella enterica* subsp. *enterica* serovar Cerro in a U.S. dairy herd. *Foodborne Pathog. Dis*. 4:449-461.

Waak, E., W. Tham, and M-L Danielsson-Tham. 2002. Prevalence and fingerprinting of *Listeria monocytogenes* strains isolated from raw whole milk in farm bulk tanks and in dairy plant receiving tanks. *Appl. Environ. Microbiol*. 68:3366-3370.

Weller, D., A. Andrus, M. Wiedmann, and H. C. den Bakker. 2015. *Listeria booriae* sp. nov. and *Listeria newyorkensis* sp. nov., from food processing environments in the USA. *Int. J. Syst. Evol. Microbiol*. 65:286-292.

Wilkin, E., R. Brouillette, J. Carver, B. Cords, A. Dhillon, S. Hall, T. Hedge, D. Kedzierski, L. Ledenbach, V. Lewandowski, M. Smith, J. Stout, and T. Stubbs. 2015. Control of *Listeria monocytogenes*: Guidance for the U.S. Dairy Industry. Innovation Center for U.S. Dairy. Available at: http://www.idfa.org/docs/default-source/resource- library/guidance-for-the-us-dairy-industry-10-19-15.pdf

Wulff, G., L. Gram, P. Ahrens, and B. F. Vogel. 2006. One group of genetically similar *Listeria monocytogenes* strains frequently dominates and persists in several fish slaughter and smokehouses. *Appl. Environ. Microbiol*. 72:4313-4322.

CHAPTER 3

CHARACTERIZATION OF PAENIBACILLUS ODORIFER *ISOLATES REVEALS PSYCHROTOLERANCE IS NOT ASSOCIATED WITH PHYLOGENY*

In Preparation for Submission to Journal of Dairy Science

ABSTRACT

Paenibacillus, a spore-forming bacterial genus, has frequently been associated with fluid milk spoilage. Using previously reported collections of dairy associated isolates we selected 101 isolates, including 58 *Paenibacillus*, 36 *Bacillus*, and 7 representing other Bacillales genera (e.g., *Solibacillus*). These isolates were initially screened for ability to grow over 21 days in Skim Milk Broth (SMB) at 6°C. While most *Paenibacillus* showed growth in SMB at 6°C, the other spore-formers generally did not show growth under these conditions. Whole genome sequencing data for the 58 *Paenibacillus* were initially used for Average Nucleotide Identity by BLAST analyses, which showed that these isolates represent 21 different species, including some groups that may represent species that have not been previously described. The plurality of *Paenibacillus* represented the species *P. odorifer*; among the 27 isolates classified into this species, 21 consistently showed growth in SMB at 6°C in the initial cold growth screen. Selected isolates representing *P. odorifer* and closely related species were further characterized for growth in SMB and Brain Heart Infusion broth at 6 and 10°C. Quality Threshold Clustering of growth data classified these isolates into 4 clusters, including 2 clusters that include isolates that showed growth in SMB at 6°C (although with different lag phase durations); all isolates showed the ability to grow in SMB at 10°C. While *P. odorifer* represented 2 well-supported clades, cold growth phenotypes did not partition into these clades; 9/13 and 5/12 isolates in clades A and B, respectively, grew in SMB at 6°C. Analyses of the genome sequences did not identify specific SNPs, genes, or Gene Ontology terms significantly associated with an isolate's ability to grow in SMB at 6°C. However, *P. odorifer* phylogenetic clades A and B showed significant

overrepresentation of 172 and 164 genes and 102 and 84 Gene Ontology terms, respectively. These results indicate that ability to grow in SMB at 6°C is not limited to a specific *P. odorifer* lineage and may represent specific adaptations that occurred independently in different subtypes. Overall, our data suggest that specific molecular assays to detect *Paenibacillus* that represent spoilage concerns in refrigerated dairy products will be challenging to develop and that comprehensive databases of spoilage associated *Paenibacillus* and their subtypes and growth phenotypes are needed to develop targeted interventions and predictive models for fluid milk spoilage.

INTRODUCTION

Paenibacillus spp. are spore-forming Bacillales with diverse morphologies and habitats (Ash et al., 1993). These bacteria are found in the natural environment and are capable of surviving and growing under a range of stress conditions (e.g., low temperature, high mineral concentrations) (Berge et al., 2002; Rivas et al., 2005). *Paenibacillus* are non-pathogenic in humans, although some isolates may have contributed to bacteremia in immune-compromised individuals (Grady et al., 2016). However, a number of *Paenibacillus* can grow at refrigeration temperatures and have been linked to food spoilage (Guinebretière et al., 2001; Ivy et al., 2012). *Paenibacillus* spp. are among the most common aerobic spore-forming bacteria isolated from HTST pasteurized milk and have been clearly shown to contribute to spoilage of HTST fluid milk (Ivy et al., 2012; Ranieri et al., 2012).

To address the issue of spoilage, it is vital to gain an improved understanding of the ability of dairy-associated *Paenibacillus* to grow in dairy products during storage at refrigeration temperatures. Studies of different organisms in the order Bacillales (which includes genera *Bacillus*, *Listeria*, and *Paenibacillus*) have identified different genes that appear to be important for the ability, of organisms in these groups, to grow at refrigeration temperatures (Chan and Wiedmann, 2009; Dsouza et al., 2014; Moreno Switt et al., 2014). For example, Moreno Switt et al. (2014) reported that cold shock proteins (**CSPs**) are commonly present in cold-adapted bacteria, including *Paenibacillus*; these proteins are likely necessary for optimal adaptation to lower temperatures (Hébraud and Potier, 1999). Dsouza et al. (2014) also reported the presence of *cspB* and *cspC* in three *P. darwinianus* isolates from Antarctica as well as in 9 temperate *Paenibacillus* spp. isolates, with 4 to 5 copies of these genes found in the *P. darwinianus* isolates (Dsouza et al., 2014). Homologs of the *Escherichia coli* cold shock protein A (CspA) have also been found in *Bacillus* spp. (Kaan et al., 2002; Budde et al., 2006; Carlin et al., 2010); in the *B. cereus* group, CspA is the predominant CSP (Carlin et al., 2010; Zhou et al., 2010). Interestingly, in the first description of *B. weihenstephanensis* (Lechner et al., 1998), the authors proposed that

this species, characterized by its ability to grow at refrigeration temperatures, can be identified using two key SNPs in *cspA* (located at *cspA* nucleotides 4 and 9) as a target; Francis et al. (1998) reported these SNPs as signatures of psychrotolerant *B. cereus* group strains compared to mesophilic strains. DEAD box helicases may also play an important role in psychrotolerance of Gram-positive bacteria (Chan and Wiedmann, 2009; Barria et al., 2013; Moreno Switt et al., 2014). These helicases improve translation by melting RNA structures and other secondary structures (Barria et al., 2013). In *E. coli*, DeaD is incorporated into the main degradosome in cold conditions, allowing for unwinding of structured RNAs (Bakersmans et al., 2012; Barria et al., 2013). Moreno Switt et al. (2014) found a DEAD box helicase present in a strain identified as *P. amylolyticus* that grows in skim milk broth (**SMB**) at 6°C. While some genes linked to cold growth in Bacillales isolates have been identified, additional studies are needed to identify specific genes related to particular spoilage organisms, such as *Paenibacillus*. The importance of specific genes for cold growth is also supported by a number of mutational analyses. For example, in *B. subtilis*, a null mutation in *des*, which encodes for the fatty acid desaturase protein Des, resulted in cold sensitivity of the mutant strain compared to the wild type when isoleucine was absent (Aguilar et al., 1998; Weber et al., 2001). In *L. monocytogenes*, a transposon insertion in lmo1078, which encodes a putative UDP-glucose pyrophosphorylase, resulted in a cold-sensitive phenotype; complementation of the mutant strain with a wild type copy of lmo1078 restored its ability to grow at low temperature (Chassaing and Auvray, 2007). Deletion mutations in CSPs have also been shown to led to impaired growth at low temperatures in *B. subtilis* (Graumann et al., 1997) and *E. coli* (Xia et al., 2001).

Understanding the relationship between psychrotolerance and *Paenibacillus* genomic and genetic characteristics has the potential to not only provide new insights into cold growth characteristics, but also may yield new genetic markers that could be used to develop rapid screening methods for detection of cold growing *Paenibacillus*. The purpose of this study hence was to screen spore-forming Bacillales for their ability to grow in SMB at 6°C, followed by an in-

depth evaluation of a selected subgroup commonly found in fluid milk (i.e., *P. odorifer*) for ability to grow at low temperatures (i.e., 6 and 10°C) in Brain Heart Infusion (BHI) broth and SMB. These phenotypic data were then combined with whole genome sequencing data to identify whether genetic or genomic markers are associated with the ability of *P. odorifer* to grow at low temperatures. SMB was chosen as a model system in our study as it provides a standardized media that mimics the composition of skim milk; SMB has previously been used to assess growth and other characteristics of spoilage organisms relevant to fluid milk (Shen et al., 2007, Ivy et al., 2012, Masiello et al., 2014, Moreno Switt et al., 2014, Trmčić et al., 2015, Masiello et al., 2016, Masiello et al., 2017). While we appreciate that it is unlikely that single genetic change is responsible for the ability of a given *Paenibacillus* strain's ability to grow at 6°C, we hypothesized that specific genetic markers (gene presence/absence, and, less likely, single nucleotide polymorphisms [SNPs]) and/or overrepresentation of specific gene sets (i.e., gene ontology [GO] terms) would be associated with ability of *Paenibacillus* strains or clonal groups to grow at 6°C in SMB.

MATERIALS AND METHODS

Strain Selection for Cold Growth Analysis

An isolate set of spore-forming Bacillales representing different subtypes commonly found in dairy associated sources was assembled for initial characterization of cold growth capabilities; subtype classification was based on *rpoB* allelic types (**AT**s) that were previously reported for all isolates (Ivy et al., 2012; Trmčić et al., 2015). Isolates were selected from a comprehensive set of 1,288 dairy associated Bacillales isolates previously described by Ivy et al. (2012) and a standard set of dairy spoilage isolates (Trmčić et al., 2015). The isolate set included 101 isolates representing seven genera, including *Paenibacillus* (58 isolates), *Bacillus* (36 isolates), *Viridibacillus* (2 isolates), *Psychrobacillus* (2 isolates), *Solibacillus* (1 isolate), *Oceanobacillus* (1 isolate), and *Lysinibacillus* (1 isolate) (Supplemental Table 3.1). The selected isolate set

represented 93 different *rpoB* ATs; eight *rpoB* ATs (ATs 1, 2, 27, 35, 46, 73, 159, and 179) were represented twice in the isolate set, including AT 1, which was the most common subtype represented among the dairy associated isolates reported by Ivy et al. (2012). Among the 101 isolates, 74 represent *rpoB* ATs that were represented at least three times in the comprehensive isolate set reported by Ivy et al. (2012), while 22 isolates were conveniently selected to represent less common *rpoB* ATs (i.e., ATs represented by 1 or 2 isolates in the data set described by Ivy et al. [2012]). An additional 5 isolates were selected to represent *rpoB* ATs that were discovered following the publication by Ivy et al. (2012); these isolates were described by Trmčić et al. (2015). All isolates represent dairy associated sources, including milk (raw, lab-pasteurized, or pasteurized) or dairy farm environments.

Screening for Growth in SMB at Refrigeration Temperature

All 101 isolates were initially screened for the ability to grow at refrigeration temperature under a single condition, SMB at 6°C, using methodology previously described by Ivy et al. (2012). Briefly, isolates were grown in brain heart infusion (**BHI**) broth at 32°C; these cultures were then used to inoculate SMB to a concentration of approximately 10^2 CFU/ml. Cultures were incubated (without aeration) at 6°C for 21 days. At days 0, 14, and 21, cultures were spiral plated using an Autoplate 5000 (Advanced Instruments, Inc., Norwood, MA) onto Standard Plate Count (**SPC**) agar using the 50 µl exponential setting. Plates were incubated at 32°C for 24-48 h and colonies were counted using a Q-Count Colony Counter (Model 530, Advanced Instruments, Inc., Norwood, MA). All isolates were tested at least twice.

Quantification of Growth at 6 and 10°C in BHI and SMB.

Twenty-five isolates representing *P. odorifer* and three isolates representing closely related species were tested for growth in SMB and BHI broth over 21 days at 6 and 10°C. Briefly, isolates were streaked, in triplicate, onto BHI agar from frozen glycerol stock. These agar plates were

incubated at 32°C for 18-24 h. One colony of each replicate was inoculated into 5 ml of BHI broth and incubated at 32°C for approximately 24 h. Cultures were diluted and spiral plated onto BHI agar using the 50 µl exponential setting in technical duplicates. After 24 h of incubation at 32°C, colonies were enumerated using the Q-Count Colony Counter. The bacterial numbers obtained were used to dilute the original cultures (which were stored at 4°C for 24 h) to allow for inoculation into fresh SMB and BHI tubes to a final concentration of approximately 10^2 CFU/ml. Inoculated SMB and BHI tubes were incubated at either 6 or 10°C (without aeration). Each triplicate was spiral plated onto two BHI agar plates at each time point (immediately after inoculation, as well as after 14 and 21 days of incubation at either 6 or 10°C). BHI agar plates were incubated at 32°C for approximately 24 h and subsequently, colonies were enumerated. Samples without any bacterial growth on either of the duplicate plates, indicating bacterial counts below the detection limit of 10 CFU/ml, were arbitrarily and conservatively counted as 1 CFU/ml.

Statistical Analyses of Cold Growth Data

To assess the effect of media and temperature on the growth of the 28 *Paenibacillus* strains tested, a linear regression was used to fit the data with the censReg package (Henningsen, 2017) in R, with "day of testing" as a categorical effect and a two-way interaction of "media" and "temperature"; below detection observations were included as left-censored observations.

Quality Threshold Clustering (**QTC**) was used to group isolates by growth patterns. For this analysis, counts were averaged for the replicates of each isolate and averages were normalized to day 0 counts. QTC was performed using the qtclust function (Scharl and Leisch, 2006) of the flexclust package (Leisch, 2006) in R and a radius of 1.5 log CFU/ml.

DNA Extraction and Whole-Genome Sequencing of **Paenibacillus** *spp.*

As *Paenibacillus* spp. are among the most common spore-forming contaminants in milk (Ivy et al., 2012; Ranieri et al., 2012), we performed whole genome sequencing (**WGS**) of 55 of

the 58 *Paenibacillus* spp. isolates included in our isolates set; WGS data for the remaining three *Paenibacillus* isolates had previously been reported (Moreno Switt et al., 2014). Additionally, three *Viridibacillus* spp. isolates were selected for WGS. DNA extraction was performed using the QIAamp DNA Mini kit (Qiagen, Valencia, CA) according to the manufacturer protocol, modified to include a 45 min lysis step in a 37°C water bath. Briefly, overnight cultures of the selected isolates were pelleted and pellets were lysed with 180 µl of 20 mg/ml lysozyme (for 45 min at 37°C). DNA was eluted twice in 50 µl of Tris-HCl (pH 8.0). DNA purity was assessed using a Nanodrop (ThermoFisher Scientific, Waltham, MA) and the concentration of double-stranded DNA (**dsDNA**) was determined using a Qubit dsDNA high sensitivity kit (ThermoFisher Scientific). The concentration of dsDNA was adjusted to 1 ng/µl and DNA samples were submitted to the Cornell University Institute of Biotechnology Genomics Facility (Ithaca, NY) for Nextera XT DNA library preparation. Samples were pooled and sequenced in two Illumina HiSeq 2500 rapid runs with 2 x 100 bp paired-end reads targeting 83x and 91x coverage, respectively.

Read Processing, Quality Control, Genome Assembly, and Annotation

Nextera XT adapters, as well as low quality bases and reads, were trimmed using default settings of Trimmomatic v0.33 (Bolger et al., 2014). The quality of the short reads was assessed using FastQC (v0.11.2) (Babraham Bioinformatics). Genomes were assembled *de novo* using SPAdes v3.6.2 with a variety of k-mer sizes (21, 33, 55, 77, 99) (Bankevich et al., 2012). QUAST was used to verify the quality of the assembled draft genomes (Gurevich et al., 2013). Average coverage was determined by mapping the reads against draft genomes using BBMap v35.49 and computing the average depth using SAMtools (Li et al., 2009). Sequence reads were submitted to SRA and assembled draft genomes were submitted to the WGS NCBI database through the prokaryotic genome annotation pipeline (Tatusova, 2016).

SNP Detection, Phylogeny Construction, and Average Nucleotide Identity by BLAST (ANIb)

Analyses

SNPs were called using kSNP3 (Gardner et al., 2015). The k-mer size of 19 was selected using Kchooser in kSNP3 (Gardner et al., 2015). A maximum likelihood (ML) tree was constructed in RAxML v8.0 (Stamatakis, 2014) using core SNPs detected by kSNP3. The ML tree was constructed using a general time-reversible (**GTR**) model with gamma-distributed sites (GAMMA). The tree was rooted by midpoint and bootstrap values were based on 1000 bootstrap repetitions. The phylogenetic tree was edited in FigTree v1.4.2. ANIb analyses were performed to delineate species boundaries of all 58 *Paenibacillus* genomes (Richter and Rosselló-Móra, 2009; Miller et al., 2016).

SNP Analysis

For the 28 isolates with quantitative data for cold growth at 6 and 10°C in SMB and BHI, we assessed associations between SNPs and the ability of an isolate to grow at 6°C in SMB using two-sided Fisher's exact tests. P-values were corrected using the False Discovery Rate approach.

Identification of Proteins Related to Psychrotolerance

Twelve Hidden Markov Models (**HMM**) were used to search for domains or proteins previously shown to be involved in psychrotolerance and growth at low temperatures (see Supplemental Table 3.2 for details); HMMs were obtained from Pfam 26.0 protein families' database (Finn et al., 2016). Searches against the whole genome sequences were performed using HMMER3 (Eddy, 2011). Day 14 and 21 normalized log CFU/ml growth data was transformed using Principle Component Analysis (princomp function in R), and the first principle component accounted for 79.6% of the variance. Plots were visualized using ggplot (Wickham, 2009). As the HMM protein family searches identified multiple hits (genes) for at least some genomes in each search, linear regression was run in R to determine whether HMM protein family matches per genome are associated with the first principle component of day 14 and 21 growth in SMB at 6°C.

OrthoMCL and Gene Ontology (GO) Term Annotation

OrthoMCL (Li et al., 2003) was used with three inflation values (5, 3, and 1.5) to find ortholog clusters (groups of orthologous genes found across multiple isolates) in the genomes representing 25 *P. odorifer* isolates and 3 isolates of closely related species with quantitative data for cold growth at 6 and 10°C. Higher inflation values result in higher ortholog cluster tightness. Briefly, ortholog clusters with 28 genomes using an inflation value of 5 were identified first. Next, ortholog clusters with 28 genomes using an inflation value of 3 were identified; excluding the ortholog clusters selected using an inflation value of 5, followed by selection of ortholog clusters with 28 genomes and an inflation value of 1.5, identifying clusters that were not previously identified using inflation values of 3 and 5. This approach was used to minimize the inclusion of more than one gene from the same genome in a given ortholog cluster. Finally, ortholog clusters with less than 28 genomes with an inflation value of 1.5 were identified, resulting in a total of 10,070 clusters.

One representative protein sequence from each cluster was selected for gene ontology annotation using Blast2GO (Conesa and Götz, 2008). Protein sequences were first searched against the SWISS-PROT database (Bairoch and Apweiler, 2000). Genes that were not assigned any gene ontology (**GO**) terms were searched against the RefSeq database (O'Leary et al., 2016). The outputs were combined and the assigned GO terms were linked to their respective ortholog clusters and to each member of the ortholog cluster.

Gene Presence/Absence Analysis and Gene Enrichment

For each ortholog cluster, counts of genomes where the gene was present or absent were computed for (i) "growers" and "non-growers", or for (ii) clade A and clade B members; for the analyses under (i), "growers" were defined as isolates that showed at least a 1 log increase in bacterial numbers at either day 14 or 21, all other isolates were designated as "non-growers". In addition, for each GO term, the numbers of present and absent genes were computed for (i)

"growers" and "non-growers", or for (ii) clade A and clade B members. Next, 2x2 tables were generated for each GO term, and two-sided Fisher's exact tests were run for each GO term. Odds ratios were computed and p-values were adjusted using the False Discovery Rate approach.

RESULTS

Initial Cold Growth Screen of Dairy Associated Bacillales Isolates Shows that Ability to Grow at 6°C in SMB is Primarily Associated with **Paenibacillus** *Isolates*

An initial screen for ability to grow in SMB at 6°C over 21 days was used to identify isolates with ability to grow under these conditions; for this initial screen, growth was defined as at least a 1 log increase in bacterial numbers at either day 14 or 21. Of the 101 total isolates, 34 consistently showed growth (mean growth of 2.58 and 4.00 log at days 14 and 21, respectively), while 55 consistently showed no growth (see Supplemental Table 3.1 for detailed data). An additional 12 isolates displayed "stochastic growth", which we defined as growth in only some of the replicates (see Supplemental Table 3.1); at least two replicates were performed for each isolate.

The fact that the 101 isolates screened for cold growth in SMB represented a range of dairy associated genera and species also allowed us to explore the distribution of ability to grow in SMB at 6°C across different taxonomic groups. With regard to the less common genera, all *Psychrobacillus* spp. (n=2), *Lysinibacillus* sp., *Oceanobacillu*s sp., and *Soliobacillus* sp. isolates (all n=1) were unable to grow in SMB at 6°C. Of the two *Viridibacillus* sp. isolates tested, one grew in SMB at 6°C, while the other isolate displayed stochastic growth. Among the 36 *Bacillus* spp. isolates tested, 35 did not grow at 6°C in SMB; the one remaining isolate (which was classified as *B. weihenstephanensis*) displayed stochastic growth at 6°C in SMB. Among the 58 *Paenibacillus* spp. isolates, 33 grew at 6°C in SMB, 15 did not grow, and 10 displayed stochastic growth.

Whole Genome Sequence Data Analyses Reveal a Number of **Paenibacillus** *Clades that may*

Represent New Species and Show a Range of* Paenibacillus *Genome Sizes that Range from 6.0 to 9.5 Mb.

As our screen identified that growth in SMB at 6°C was most common among *Paenibacillus* isolates, we pursued further characterization of these isolates through WGS (see Table 3.1 for details and accession numbers). Initial ANIb analyses of the genome sequences for the 58 *Paenibacillus* isolates characterized here along with 16 sequences representing relevant type strains (i.e., *P. odorifer* DSM 15391^T, *P. graminis* DSM 15220^T, *P. riograndensis* SBR5^T, *P. sonchi* X19-5^T, *P. borealis* DSM 13188^T, *P. stellifer* DSM 14472^T, *P. sabinae* T27^T, *P. forsythiae* T98^T, *P. durus* DSM 1735^T, *P. zanthoxyli* JH29^T, *P. glucanolynicus* DSM 5162^T, *P. macerans* ATCC 8244^T, *P. massiliensis* 2301065^T, *P. peoriae* KCTC 3763^T, *P. pabuli* NBRC 13638^T, and *P. wynnii* LMG 22176^T) were performed to verify genus and species assignment of the dairy isolates characterized here, which was based on *rpoB* and 16S rDNA sequence data as described by Ivy et al. (2012). An initial 16S rDNA based phylogeny was used to select those type strains with genomes that were closely related to the 58 *Paenibacillus* isolates characterized here. Using a cut-off of 95% ANIb for classification into the same species (Chan et al., 2012), the ANIb analyses indicated that the 58 dairy *Paenibacillus* isolates characterized represent 21 different *Paenibacillus* species with 27, 2, and 1 isolates classified into the species *P. odorifer*, *P. glucanolyticus*, and *P. macerans*, respectively (Figure 3.1). The remaining 28 isolates represented 18 additional *Paenibacillus* spp. that could not be reliably classified to the species level, as they did not show similarities of ≥ 95% with any of the type strain genomes included in our analyses. Interestingly, two *Paenibacillus* type strains used in the ANIb analyses (*P. riograndensis* SBR5^T and *P. sonchi* X19-5^T) show an ANIb similarity score of 98% and hence do not meet the species cut-off of < 95%.

Genome sizes among the whole set of 58 *Paenibacillus* isolates ranged from 6.0 to 9.5 Mb. *P. odorifer*, which represented the most common species found among dairy related isolates characterized here, showed a range of assembly sizes from 6.8 to 7.5 (based on 27 isolates); the *P.*

odorifer type strain (DSM 15391) was reported to show a genome size of 6.8 Mb (see (https://www.ncbi.nlm.nih.gov/assembly/GCF_000758725.1). By comparison, the three *Viridibacillus* isolates sequenced showed assembly sizes between 4.4 and 4.5 Mb. Nucleotide coverages for the 58 *Paenibacillus* and 3 *Viridibacillus* genomes ranged from 35 x to 175 x; all genomes that have been newly sequenced here have been annotated through NCBI's Prokaryotic Annotation Pipeline and are available for future studies and analyses.

TABLE 3.1: Detailed information on the 58 *Paenibacillus* isolates and 3 *Viridibacillus* isolates characterized by whole genome sequencing

Strain	Genus	species	*rpoB* AT	Source	Number of contigs	Number of genes[a]	Assembly Size	Nucleotide Coverage	SRA Accession	WGS Accession
FSL A5-0030	*Paenibacillus*	sp.	179	pasteurized milk	74	5354	6 Mb	94	SRR4434617	MRTC00000000
FSL A5-0031	*Paenibacillus*	sp.	511	pasteurized milk	87	6862	7.8 Mb	81	SRR4434616	MRTD00000000
FSL F4-0077	*Paenibacillus*	*odorifer*	2	pasteurized milk	25	6061	6.9 Mb	102	SRR4434615	MPVO00000000
FSL F4-0085	*Paenibacillus*	*odorifer*	4	pasteurized milk	106	5898	6.8 Mb	80	SRR4242611	MPTE00000000
FSL F4-0087	*Paenibacillus*	sp.	5	pasteurized milk	33	6142	6.9 Mb	70	SRR4434614	MRTE00000000
FSL F4-0100	*Paenibacillus*	sp.	8	pasteurized milk	80	6940	7.7 Mb	72	SRR4434613	MRTF00000000
FSL F4-0126	*Paenibacillus*	*odorifer*	13	pasteurized milk	128	6124	6.9 Mb	71	SRR4242620	MPTU00000000
FSL F4-0134	*Paenibacillus*	*odorifer*	16	pasteurized milk	102	6075	6.9 Mb	72	SRR4242605	MPTJ00000000
FSL F4-0152	*Paenibacillus*	*odorifer*	19	pasteurized milk	130	6311	7 Mb	105	SRR4242594	MKQK00000000
FSL F4-0242	*Paenibacillus*	*odorifer*	25	pasteurized milk	173	6744	7.5 Mb	65	SRR4242600	MPTN00000000
FSL F4-0260	*Paenibacillus*	sp.	29	pasteurized milk	62	6148	7 Mb	74	SRR4434612	MRTG00000000
FSL H3-0280	*Paenibacillus*	*odorifer*	27	raw milk	142	6398	7.1 Mb	77	SRR4242617	MKQO00000000
FSL H3-0287	*Paenibacillus*	*odorifer*	2	raw milk	94	6057	7 Mb	73	SRR4242603	MPTK00000000
FSL H3-0305	*Paenibacillus*	*odorifer*	38	pasteurized milk	59	6221	7 Mb	153	SRR4434611	MPVM00000000
FSL H3-0464	*Paenibacillus*	*odorifer*	46	pasteurized milk	111	6325	7.2 Mb	101	SRR4242597	MPTQ00000000
FSL H3-0465	*Paenibacillus*	*odorifer*	50	pasteurized milk	105	6104	7 Mb	96	SRR4242622	MPTS00000000
FSL H7-0326	*Paenibacillus*	sp.	57	pasteurized milk	27	5741	6 Mb	129	SRR4434610	MPVN00000000
FSL H7-0331	*Paenibacillus*	sp.	58	pasteurized milk	171	8450	9.5 Mb	69	SRR4434619	MRTH00000000
FSL H7-0433	*Paenibacillus*	*odorifer*	36	pasteurized milk	1,049	7223	7.5 Mb	54	SRR4242604	MPVP00000000
FSL H7-0443	*Paenibacillus*	sp.	30	pasteurized milk	70	6018	6.9 Mb	58	SRR4242613	MPTM00000000
FSL H7-0596	*Viridibacillus*	sp.	73	pasteurized milk	31	4430	4.5 Mb	155	SRR4434618	MSPV00000000
FSL H7-0604	*Paenibacillus*	*odorifer*	74	pasteurized milk	111	6494	7.3 Mb	83	SRR4242618	MKQP00000000
FSL H7-0692	*Paenibacillus*	sp.	80	pasteurized milk	36	6170	7 Mb	86	SRR4434623	MRTI00000000
FSL H7-0694	*Paenibacillus*	*odorifer*	35	pasteurized milk	115	5942	6.8 Mb	75	SRR4242602	MPTL00000000
FSL H7-0710	*Paenibacillus*	sp.	81	pasteurized milk	94	5958	6.7 Mb	104	SRR4242614	MPTC00000000
FSL H7-0713	*Paenibacillus*	*odorifer*	33	pasteurized milk	89	5987	6.8 Mb	75	SRR4242601	MPTM00000000

FSL H7-0718	*Paenibacillus*	*odorifer*	32	pasteurized milk	106	6085	7 Mb	77	SRR4242598	MPTP00000000
FSL H7-0744	*Paenibacillus*	sp.	41	raw milk	192	6882	7.7 Mb	55	SRR4242615	MPTB00000000
FSL H7-0918	*Paenibacillus*	*odorifer*	88	pasteurized milk	106	6490	7.3 Mb	89	SRR4242599	MPTO00000000
FSL H8-0069	*Paenibacillus*	*odorifer*	93	pasteurized milk	118	6092	6.9 Mb	102	SRR4242610	MPTF00000000
FSL H8-0123	*Viridibacillus*	sp.	73	pasteurized milk	37	4383	4.4 Mb	175	SRR4434622	MSPU00000000
FSL H8-0147	*Paenibacillus*	*odorifer*	40	pasteurized milk	114	6216	7 Mb	66	SRR4242608	MPTH00000000
FSL H8-0175	*Paenibacillus*	*odorifer*	107	raw milk	108	6199	7.1 Mb	89	SRR4242607	MPTI00000000
FSL H8-0237	*Paenibacillus*	*odorifer*	15	pasteurized milk	115	6526	7.3 Mb	103	SRR4242619	MPTV00000000
FSL H8-0246	*Paenibacillus*	sp.	108	pasteurized milk	57	5993	6.7 Mb	121	SRR4434621	MRTJ00000000
FSL H8-0259	*Paenibacillus*	sp.	109	pasteurized milk	66	7077	8.2 Mb	74	SRR4434620	MRTL00000000
FSL H8-0548	*Paenibacillus*	sp.	156	water from hose, milking parlour	181	6168	7.2 Mb	90	SRR4434627	MRTK00000000
FSL H8-0551	*Paenibacillus*	sp.	157	raw milk	72	5470	6.1 Mb	102	SRR4434626	MRTM00000000
FSL J3-0120	*Paenibacillus*	sp.	340	pasteurized milk	90	5372	6 Mb	93	SRR4434625	MRTN00000000
FSL J3-0122	*Paenibacillus*	sp.	23	pasteurized milk	55	6113	7 Mb	87	SRR4434624	MRTO00000000
FSL J3-0153	*Paenibacillus*	*odorifer*	46	pasteurized milk	119	5925	6.8 Mb	77	SRR4242596	MPTR00000000
FSL J3-0155	*Paenibacillus*	*odorifer*	35	pasteurized milk	129	6220	7.1 Mb	60	SRR4242609	MPTG00000000
FSL J3-0159	*Paenibacillus*	*odorifer*	7	pasteurized milk	105	6131	7 Mb	83	SRR4242616	MKQN00000000
FSL R5-0213	*Viridibacillus*	*arenosi*	17	pasteurized milk	44	4333	4.4 Mb	137	SRR4434629	MSPW00000000
FSL R5-0378	*Paenibacillus*	sp.	214	pasteurized milk	77	7167	7.8 Mb	81	SRR4434628	MRTP00000000
FSL R5-0490	*Paenibacillus*	sp.	93	pasteurized milk	75	4699	4.7 Mb	145	SRR4434636	MRTQ00000000
FSL R5-0527	*Paenibacillus*	*macerans*	238	pasteurized milk	2,195	8988	8.4 Mb	77	SRR4434637	MRTR00000000
FSL R5-0636	*Paenibacillus*	*odorifer*	180	pasteurized milk	80	6264	7 Mb	76	SRR4242595	MKQL00000000
FSL R5-0765	*Paenibacillus*	sp.	261	pasteurized milk	85	5942	6.9 Mb	94	SRR4434634	MRTS00000000
FSL R5-0808	*Paenibacillus*	*glucanolyticus*	159	pasteurized milk	NA*	NA*	6.4 Mb	40	NA*	ASPT00000000
FSL R5-0817	*Paenibacillus*	*glucanolyticus*	159	pasteurized milk	118	6573	7.1 Mb	77	SRR4434635	MRTT00000000
FSL R5-0883	*Paenibacillus*	*odorifer*	27	pasteurized milk	143	6434	7.2 Mb	77	SRR4242606	MKQM00000000
FSL R5-0923	*Paenibacillus*	sp.	167	pasteurized milk	80	6166	7.1 Mb	91	SRR4242612	MPTD00000000
FSL R5-0937	*Paenibacillus*	*odorifer*	21	pasteurized milk	141	6723	7.4 Mb	82	SRR4242621	MPTT00000000
FSL R7-0131	*Paenibacillus*	sp.	179	pasteurized milk	71	5415	6.1 Mb	93	SRR4434632	MRTU00000000
FSL R7-0269	*Paenibacillus*	sp.	163	pasteurized milk	161	NA*	7.5 Mb	40	NA*	ASPS00000000
FSL R7-0273	*Paenibacillus*	sp.	193	pasteurized milk	85	6231	7.2 Mb	57	SRR4434633	MRTY00000000

FSL R7-0277	*Paenibacillus*	sp.	45	pasteurized milk	122	NA*	7.6 Mb	40	NA*	ASPX00000000
FSL R7-0321	*Paenibacillus*	sp.	199	pasteurized milk	58	5405	6.1 Mb	86	SRR4434630	MRTV00000000
FSL R7-0333	*Paenibacillus*	sp.	201	pasteurized milk	132	6763	7.7 Mb	35	SRR4434631	MRTW00000000
FSL R7-0337	*Paenibacillus*	sp.	202	pasteurized milk	98	6751	7.6 Mb	83	SRR4434638	MRTX00000000

* Sequences of these isolates are previously published (Moreno Switt et al., 2014)
[a] Identified by NCBI's Prokaryotic Annotation Pipeline

ANIb Analyses Reveal **Paenibacillus** *Species with Consistent Ability to Grow at 6°C in SMB as well as Other Species that Include Isolates With and Without Ability to Grow Under these Conditions.*

Mapping of cold growth screening results onto the ANIb cladogram identified some species where isolates showed consistent growth phenotypes, while other species included isolates that showed distinctly different growth phenotypes (Figure 3.1). For example, 5 isolates that showed > 95% ANIb similarity to each other and hence represent a single species (denoted as *Paenibacillus sp.* 14 in Figure 3.1) all displayed growth in SMB at 6°C. Similarly, both *P. glucanolyticus* isolates were unable to grow in SMB at 6°C. Interestingly, *P. odorifer* showed considerable variation in growth abilities; 21/27 isolates grew in SMB at 6°C, 1 isolate did not grow, and 5 isolates showed stochastic growth under these conditions.

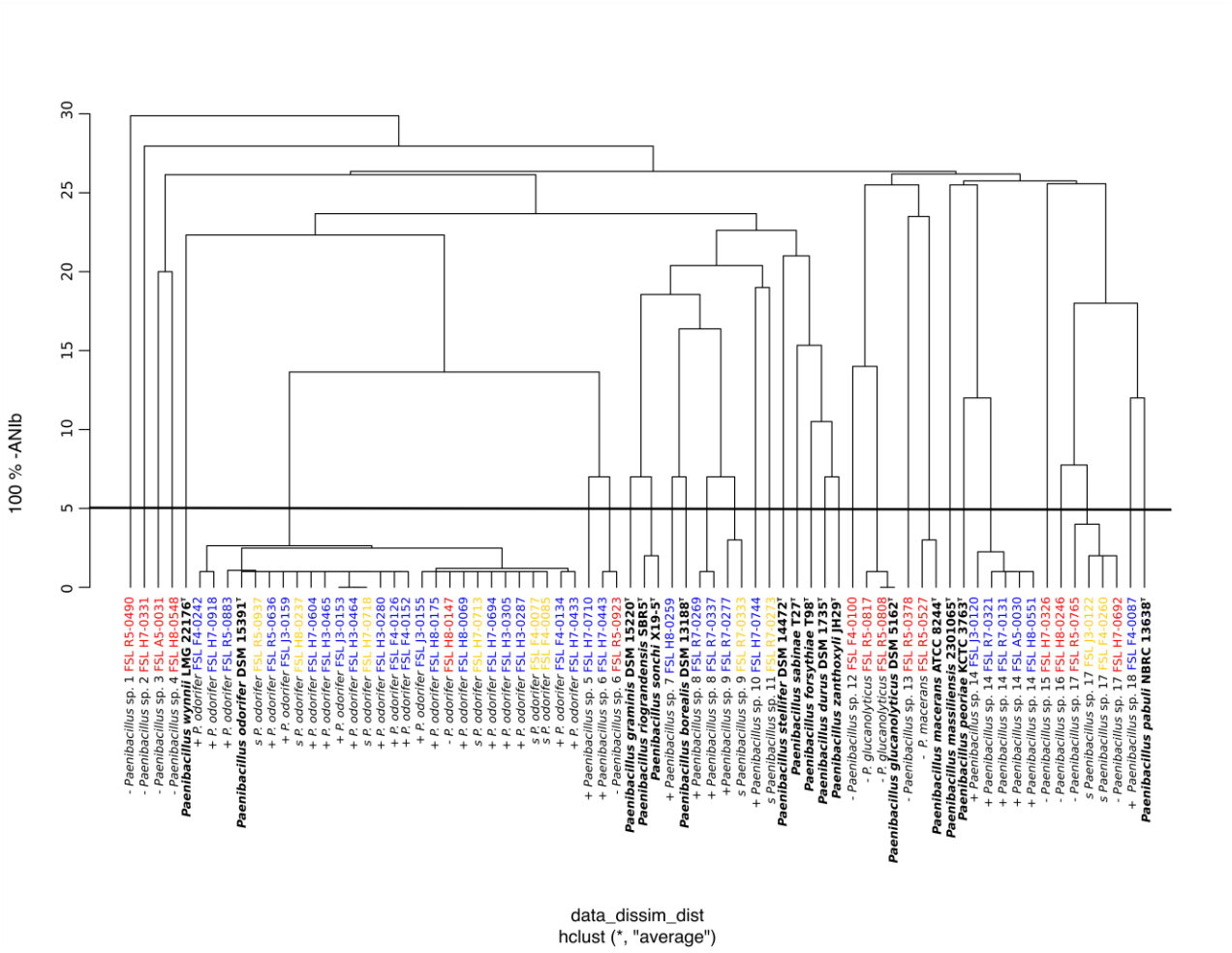


FIGURE 3.1: Average nucleotide identity by BLAST (ANIb) cladogram for 58 *Paenibacillus* isolates. The cladogram was built, in R, based on percentage average nucleotide identity by BLAST (ANIb) using the unweighted pair group method with arithmetic mean (UPGMA). The vertical bar denotes 100% - ANIb and the horizontal line denotes a cut-off of 100% - 95% ANIb; this cut-off can be used to define different bacterial species and correlates to a 70% traditional DNA-DNA hybridization value cut-off (Chan et al., 2012). The 16 type strains are shown in bold type. Isolates marked in blue and with "+" showed growth at 6°C in SMB (defined as all replicates showing at least 1 log higher counts at either day 14 or day 21 than on day 0), those marked in red and with "-" did not show growth under these conditions. Isolates marked in yellow and with "s" showed stochastic growth capabilities in SMB at 6°C (at least one replicate grew while others did not).

Based on QTC, Isolates Representing* P. odorifer *and Closely Related Species Represent Four Major "Cold Growth Groups"

Based on the screening data reported above, we performed further in-depth characterization of 28 isolates (25 isolates representing *P. odorifer* and 3 isolates representing closely related unnamed species [*Paenibacillus* spp. 5 and 6; see Figure 3.1]) for growth under four different conditions, including in rich media (i.e., BHI at 6 and 10°C) and in SMB at 6 and 10°C (representing conditions typical for fluid milk) (see Supplemental Table 3.3 for detailed results). Each isolate was tested in triplicate. All 28 tested isolates were able to grow in both BHI and SMB at 10°C over 21 days, with growth defined as at least a mean 1 log increase at either day 14 or 21. Mean growth for all isolates was 4.8 log in BHI at both days 14 and 21, and 4.4 and 4.7 log in SMB at day 14 and 21, respectively. Similarly, all 28 isolates grew at least 1 log (by either day 14 or 21) in BHI broth at 6°C (mean growth of 3.9 and 4.8 log, respectively). By comparison, only some isolates showed evidence for growth in SMB at 6°C over 21 days; based on their ability to grow in SMB at 6°C, isolates could be grouped into three categories, including (i) 16 isolates that showed at least a mean 1 log increase in bacterial numbers at either day 14 or 21 (designated as category "Growth"; see Table 3.2); (ii) 10 isolates that showed at least 1 log decline at either day 14 or 21 and did not show more than 1 log growth at either time point (category "Die Off"); or (iii) 2 isolates that showed neither more than 1 log growth nor die off at either time point (category "No Growth").

TABLE 3.2**:** Cluster and numerical categorization of 25 *P. odorifer* and 3 closely related isolates (representing *Paenibacillus* spp. 5 and 6) based on growth patterns at 6°C in Skim Milk Broth (SMB) over 21 days.

Isolate ID (FSL)	*rpoB* AT	Year collected	Relative Log CFU at[b] Day 14	Relative Log CFU at[b] Day 21	Cluster[c]	Numerical category[e]
CLADE A (*P. odorifer*)[a]						
H7-0718	32	2005	-0.76	1.2	1	Growth
H8-0237	15	2005	1.3	1.6	1	Growth
J3-0153	46	2012	-1.5	1.9	1[d]	Growth
R5-0883	27	2006	1.4	2.9	1	Growth
R5-0937	21	2006	-1.0	5.0	1[d]	Growth
F4-0152	19	2002	-0.10	-2.1	2	Die off
H3-0465	50	2005	-0.79	1.9	2	Die off
H7-0604	74	2005	-1.4	-1.4	2	Die off
F4-0126	13	2002	3.4	1.4	3	Growth
H3-0280	27	2005	3.1	4.0	3	Growth
J3-0159	7	2012	3.0	3.8	3	Growth
R5-0636	180	2006	3.4	3.9	3	Growth
H3-0464	46	2005	0.38	-0.13	4	No Growth
CLADE B (*P. odorifer)*						
H7-0694	35	2005	0.35	2.4	1	Growth
H7-0713	33	2005	-1.0	2.8	1	Growth
H8-0175	107	2005	0.44	2.7	1	Growth
J3-0155	35	2012	0.21	3.4	1	Growth
F4-0134	16	2002	-1.9	-3.4	2[d]	Die off
F4-0242	25	2003	-2.1	-0.47	2	Die off
H7-0433	36	2005	-1.8	-2.6	2	Die off
H7-0918	88	2005	-2.0	-2.0	2	Die off
H8-0069	98	2005	-1.7	0.24	2	Die off
H8-0147	40	2005	-2.0	-0.61	2	Die off
F4-0085	4	2002	0.91	-1.6	4	Die off
H3-0287	2	2005	1.3	-0.04	4	Growth
CLADE C (*Paenibacillus* species 6)						
H7-0443	30	2005	-0.54	1.0	1	Growth
R5-0923	167	2006	-0.46	0.53	4	No Growth
CLADE D (*Paenibacillus* species 5)						
H7-0710	81	2005	0.94	3.2	1	Growth

[a] Clade designations correspond to clades shown in Figure 3.3

[b] log counts have been normalized to bacterial numbers at day 0

[c] Clusters were assigned using Quality Threshold Clustering (QTC) with a radius of 1.5 log CFU/ml (see Figure 3.2).

[d] These isolates were not assigned into clusters based on QTC, but were manually added to the indicated QTC clusters, which contained isolates with similar growth patterns.

[e] Numerical groups were assigned using the following guidelines. (i) Growth: showed at least a 1 log increase in bacterial numbers at either day 14 or 21; (ii) Die off: showed at least a 1 log decline in bacterial numbers at either day 14 or 21 and did not show at least a 1 log increase at either time point; (iii) No Growth: did not show at least a 1 log increase or decline at day 14 or 21.

To provide for additional classification of isolates based on growth patterns in SMB at 6°C, we performed QTC with a radius of 1.5 log CFU/ml on the mean day 14 and 21 counts in SMB incubated at 6°C. Clustering based on the 6°C SMB data revealed four QTC clusters in which 25/28 isolates were grouped (Figure 3.2). The nine isolates in QTC Cluster 1 (see Table 3.2 and Figure 3.2) can be described as showing growth with an elongated lag phase, as compared to QTC Cluster 3, which shows growth with a short lag phase. The eight isolates in QTC Cluster 2 can be described as showing die off at 6°C in SMB; all isolates had counts at days 14 and/or 21 that were lower than the count at day initial (1.53 and 1.58 log mean die off at days 14 and 21, respectively). The four isolates in QTC Cluster 3 can be described as showing growth with a shorter lag phase, as compared to QTC Cluster 1; these isolates grew at least 3 log after 14 days. QTC Cluster 4 includes four isolates that can be described as showing no or limited growth; this cluster included the two isolates in the "no growth" group (see above) as well as one isolate (FSL H3-0287) that grew 1.3 log by day 14, but showed day 21 bacterial counts close to the day initial starting numbers and one isolate (FSL F4-0085) that showed a 1.6 log die off by day 21 (Table 3.2). Three isolates (FSL F4-0134, FSL J3-0153, and FSL R5-0937) were not grouped into any of these four QTC clusters (Figure 3.2A). For the analyses shown below that used QTC clusters, isolates FSL J3-0153 and FSL R5-0937 were added to QTC cluster 1 as they showed growth comparable to other QTC cluster 1 isolates (see Table 3.2 and Figure 3.2B). Similarly, isolate FSL F4-0134 (which showed more die off by day 21 than any other isolate; see Table 3.2 and Figure 3.2B) was added to QTC cluster 2.

To quantitatively evaluate growth under the different conditions, we also performed linear regression on growth at day 14 and 21. Consistent with the observations detailed above, these analyses showed that isolates grew to lower numbers in SMB at 6°C, as compared to the other growth conditions. As estimated by the linear regression, bacterial numbers in SMB at 6°C are 2.65 log lower, on average, compared to SMB at 10°C. Bacterial numbers in SMB at 6°C are 2.59 log and 2.83 log lower, on average, compared to BHI at 6°C and BHI at 10°C, respectively (see

Table 3.3 for all comparisons).

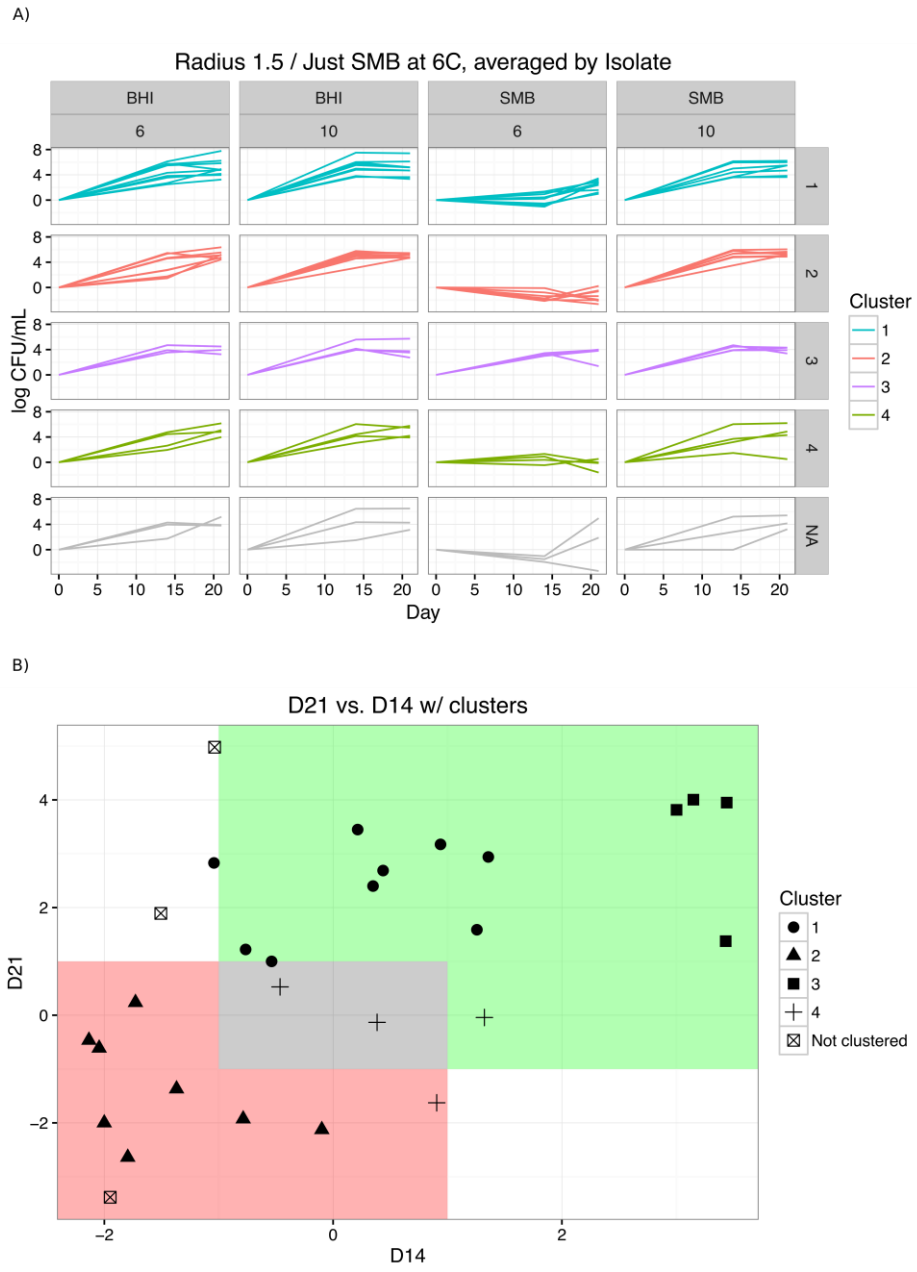


FIGURE 3.2: Growth patterns in SMB at 6°C for 25 *P. odorifer* isolates and three isolates representing closely related species. (A) Isolates were grouped based on growth in SMB at 6°C using Quality Threshold Clustering with a radius of 1.5 log CFU/ml. The three replicates were averaged for each isolate and clustered according to their growth at 6°C in SMB. Cluster color

codes correspond with the growth pattern. Isolates in QTC clusters 1 and 3 are characterized by growth at in SMB with elongated and short lag phase, respectively and are marked in blue and purple, isolates in QTC cluster 2 (characterized by "die off") are marked in red, and isolates in QTC cluster 4 (characterized as showing no or limited growth) are marked in green. The bottom panel (with growth pattern shown in gray) includes three isolates that were not classfied into any of the four QTC clusters. (B) Growth (relative to day 0) at each time point was averaged for each of the 28 isolates (three replicates) and plotted for day 14 on the X-axis and for day 21 on the Y-axis. Each Quality Threshold Cluster group is represented by different shape (e.g., triangle, square). The area shaded in green represents the category "growth", whereas the area shaded in red represents "die off" and the gray area represents "no growth" (consistent with the numerical catgories in Table 3.1; the three isolates that do not fall into any of the three colored areas also represent the numerical category "Growth" in Table 3.1 as these isolates showed > 1 log growth at day 21). Importantly, in instances where the bacterial count was below the limit of detection (i.e., < 10 CFU/ml), we arbitrarily and conservatively set this reduction at zero log (i.e., 1 CFU/ml).

Table 3.3: Regression output from quantification of growth at 6 and 10°C in BHI and SMB for 28 *Paenibacillus* isolates (25 *P. odorifer* and 3 closely related isolates representing unnamed *Paenibacillus* spp. 5 and 6).

Conditions	Contrast: Log CFU/ml change (p-value)[a]
BHI 6°C - BHI 10°C	-0.24 (0.142)
BHI 6°C - SMB 6°C	2.59 (<0.001)
BHI 6°C - SMB 10°C	0.06 (0.730)
BHI 10°C - SMB 6°C	2.83 (<0.001)
BHI 10°C - SMB 10°C	0.18 (0.263)
SMB 6°C - SMB 10°C	-2.65 (<0.001)

[a] These values represent the model estimate for growth at the first condition relative to the growth at the second condition; for example a value of 2.59 for "BHI 6°C - SMB 6°C" indicates 2.59 log higher bacterial numbers in BHI at 6°C relative to SMB at 6°C (at both t=14 and t=21 days)

WGS-based Phylogeny Revealed Two Major Clades Among **P. odorifer** *Isolates, But No Clear Clustering of Cold Growers*

Based on the variability of growth in SMB at 6°C, we decided to further explore genetic or genomic markers associated with cold growth in SMB for *P. odorifer* and closely related species (i.e., unnamed *Paenibacillus* spp. 5 and 6). As a first step, we constructed a ML core genome SNP phylogeny of the 28 isolates that were characterized in depth for growth in SMB and BHI at 6 and 10°C. The phylogenetic tree constructed from the 19,990 core SNPs identified among these 28 isolates resulted in four well-supported clades, each supported by bootstrap values of 100 (Figure 3.3). These clades were consistent with the ANIb cladogram (see Figure 3.1). Specifically, phylogenetic clade C, which contains FSL H7-0443 and FSL R5-0923 (Figure 3.3), represents unnamed *Paenibacillus* sp. 6 (Figure 3.1), while clade D, which contains a single isolate (FSL H7-0710), represents unnamed *Paenibacillus* sp. 5 (Figure 3.1). The remaining 25 isolates (representing 22 *rpoB* ATs) represent two phylogenetic clades (A and B), which both represent *P. odorifer*; these two clades are also apparent in the ANIb cladogram (Figure 3.1).

Overall, 13 and 12 of the *P. odorifer* dairy isolates with in-depth growth characterization were classified into clades A and B. Mapping of the four QTC clusters based on growth in SMB at 6°C onto the core SNP phylogeny (Figure 3.3), showed that all four isolates in QTC cluster 3 ("growth with short lag phase") grouped into clade A, while isolates in the other 3 QTC clusters were found across both clades A and B. Statistical analysis of distribution of the 4 QTC clusters among the two *P. odorifer* clades (Fisher's exact test performed on a contingency table), however, showed no significant association between clades and QTC clusters, possibly due to the low power of this analysis.

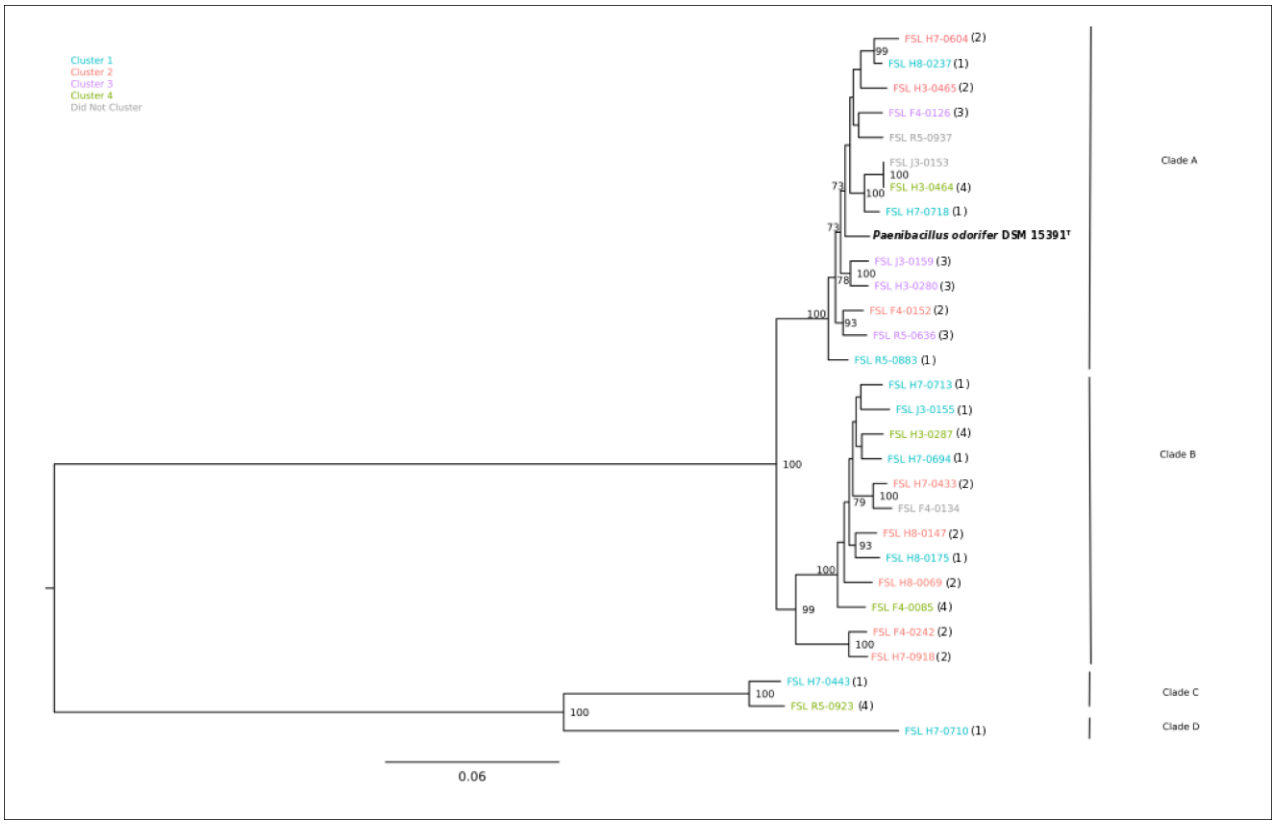


FIGURE 3.3: Phylogenetic tree constructed from the core SNPs of identified in the genome of 25 *P. odorifer* isolates and 3 isolates representing closely related species (*Paenibacillus* spp. 5 and 6). The maximum likelihood tree was constructed using a general time-reversible model with gamma-distributed sites and 1,000 bootstrap repetitions. Only bootstrap values ≥ 70 are shown on the tree. The isolates group into four clades, *P. odorifer* clades A and B, and closely related clades C and D, representing unnamed *Paenibacillus* spp. 6 and 5, respectively. QTC clusters representing isolates with similar growth patterns in SMB at 6°C (see Figure 3.2A) are mapped onto the phylogenetic tree, using the same colors for each cluster as shown in Figure 3.2A; hence isolates in QTC clusters 1 and 3 (characterized by growth in SMB at 6°C with elongated and short lag phase, respectively) are marked in blue and purple, isolates in QTC cluster 2 (characterized by "die off") are marked in red, and isolates in QTC cluster 4 (characterized as showing no or limited growth) are marked in green. The cluster number in parentheses next to each isolate name indicates which of the four cluster groups the respective isolate is associated with.

No Ortholog Clusters, GO Terms, SNPs, or HMM Protein Families were Significantly Associated with **P. odorifer** *Cold Growth in SMB*

As a next step to explore the genetic basis of cold growth in *P. odorifer*, and to identify potential genomic markers for growth in SMB at 6°C, we identified orthologs found across some or all of the 28 isolates with quantified growth data in SMB and BHI broth at 6 and 10°C, and subsequently tested for association between ortholog presence/absence and ability of isolates to grow in SMB at 6°C. For these analyses, isolates were classified as either "growers" or "non-growers" in SMB at 6°C; while "growers" were defined as isolates that showed as at least a 1 log increase in bacterial numbers at either day 14 or 21 (numerical category "growth"; see Table 3.2), all other isolates were designated as "non-growers" (numerical categories "no growth" or "die off"; see Table 3.2); this approach was chosen as only two isolates fell into the category "no growth." Among the 10,070 OrthoMCL clusters (conceptually representing 10,070 different genes), 6,112 clusters showed variable presence (meaning they were absent from at least one of the genomes). None of these clusters were significantly associated with isolate classification as "grower" or "non-grower" at 6°C in SMB. OrthoMCL clusters were also assigned into GO terms; we also found no significant associations between isolate classification as "grower" or "non-grower" and GO terms. While we also hypothesized that specific SNPs may be associated with ability to grow in SMB at 6°C, we did not identify any SNPs that were significantly overrepresented among isolates classified as "growers" or "non-growers".

HMM searches for proteins and domains associated with psychrotolerance, identified hits in the at least one genome for 10 of these HMMs, including (i) one HMM (CapC; PF14102.5) that identified matches in 8 genomes (with 1 or 2 matches per genome) and (ii) 9 HMMs that identified matches in all 28 genomes (PF00270.28, PF00768.19, PF00313.21, PF00487.23, PF00226.30, PF14169.5, PF03405.13, PF06772.10, PF00154.20; see Supplemental Table 3.4). In all cases, each hit in a given genome represented a different gene, indicating that these searches identified multiple genes rather than a single gene that contained multiple copies of a given domain. The

HMM for DEAD (PF00270.28) identified between 40 and 49 matches in a given genome, while none of the other HMMs identified more than 10 hits in a given genome (see Supplemental Table 3.4 for details). Statistical analyses showed no significant association between the number of hits with a given HMM and growth in SMB at 6°C, suggesting the number of "psychrotolerance" genes identified by HMM was not associated with ability to grow in SMB at 6°C.

A Number of Ortholog Clusters and GO Terms were Associated with the Two Major **P. odorifer** *Clades*

Since we did not identify any specific ortholog clusters or GO terms that were associated with isolate classification as "grower" or "non-grower", we further explored whether specific ortholog clusters or GO terms were significantly associated with the two major and well supported *P. odorifer* clades (A and B). We hypothesize that these two clades may have adapted to different ecological niches and that ortholog clusters or GO terms associated with these clades may provide initial insights into the ecological niches the isolates in these clades are associated with. Overall, 172 and 164 ortholog clusters and 102 and 84 GO terms were found to be significantly overrepresented in clade A and clade B isolates, respectively.

The 172 ortholog clusters overrepresented in clade A included 116 clusters annotated as hypothetical proteins. In addition, six clusters annotated as encoding proteins with nitrate and nitrite reductase related functions were found in all clade A isolates, but none of the clade B isolates. This is consistent with the observation that related GO terms (e.g., nitrate reductase (NAD(P)H), nitrogen compound metabolic process, and nitrate reductase complex, see Supplemental Tables 3.5-8) were also overrepresented in clade A. Similarly, three clusters annotated as encoding MerR family transcriptional regulators and three clusters annotated as encoding proteins with molybdenum-associated functions were identified as significantly overrepresented in clade A (Supplemental Table 3.5). Likewise, GO terms related to these processes (e.g., molybdopterin synthase activity, molybdopterin cofactor metabolic process, see

Supplemental Table 3.7) were also overrepresented in clade A.

Clade B included 164 overrepresented ortholog clusters and 84 overrepresented GO terms. Like in clade A, many (76) of the clusters overrepresented in clade B represent hypothetical proteins. In addition, six clusters of AraC family transcriptional regulators, two clusters of two-component system histidine kinases, and four clusters annotated as encoding proteins with glycoside hydrolase-associated functions were overrepresented among clade B isolates (see Supplemental Table 3.6). GO terms related to these functions (e.g., hydrolase activity, see Supplemental Table 3.8) were also overrepresented in clade B.

DISCUSSION

Spoilage of fluid milk products caused by bacteria in the order Bacillales is increasingly recognized as a major challenge for the dairy industry (Moreno Switt et al., 2014). A wide range of genera and species in the order Bacillales (e.g., *Paenibacillus, Bacillus, Listeria*, *Psychrobacillus*, *Viridibacillus*, and *Solibacillus*) have been isolated from fluid milk and from dairy farm and dairy processing environments (Ivy et al., 2012; Miller et al., 2015; Beno et al., 2016) and many of them (except, for example, *Listeria*) have been linked to spoilage of HTST-pasteurized fluid milk (Trmčić et al., 2015). While some studies (Stenfors Arnesen et al., 2007; Ivy et al., 2012) that have described isolation and initial characterization of organisms in this group from dairy sources have provided initial insight into the types of aerobic spore-forming Bacillales associated with dairy spoilage, in-depth characterization of organisms associated with and responsible for dairy spoilage is still largely missing. We thus screened an initial set of 101 Bacillales isolates for ability to grow in SMB at 6°C followed by more detailed cold growth characterization and in-depth genomic characterization of a subset of isolates focusing on the genus *Paenibacillus*, with a specific focus on *P. odorifer*, which was found to include the largest number of dairy associated isolates able to grow at low temperatures. Overall, our data show that dairy associated Bacillales isolates represent extensive taxonomic diversity as well as considerable

diversity with regard to ability to grow at refrigeration temperatures. Importantly, isolates with ability to grow at low temperatures in dairy relevant environments (i.e., SMB) represent a range of orders, genera, and subtypes and often do not represent coherent taxa or clades with clearly identifiable genetic markers for ability to grow at low temperatures in SMB. Thus, development of detection and control strategies for psychrotolerant spore-formers in the fluid milk value chain are likely to represent a considerable challenge. Importantly, our study reports cold growth data for a large and diverse set of isolates, which will be useful for future studies and modeling efforts that assess the spoilage risk associated with different Bacillales species and subtypes.

Dairy Associated Bacillales Include Diverse Groups of Isolates with Varying Growth Capabilities

An initial screen of 101 Bacillales isolates from dairy associated environments and predominantly representing the genera *Bacillus* and *Paenibacillus,* showed that 1/36 *Bacillus* isolates showed stochastic growth in SMB at 6°C, while 33/58 *Paenibacillus* were able to grow and 10/58 showed stochastic growth under these conditions. These findings are similar to previous studies, which also reported that *Paenibacillus* dairy isolates were more likely to grow at 6°C than *Bacillus* spp. For example, a study that evaluated *Bacillus* and *Paenibacillus* isolates representing common *rpoB* ATs for their ability to grow in SMB at 6°C (Ivy et al., 2012) reported that many *Paenibacillus* isolates were able to grow in these conditions, while few *Bacillus* were able to grow in these conditions. Consistent with our data reported here, this previous study (Ivy et al., 2012) also found some *Paenibacillus* isolates either did not grow or had limited growth in SMB at 6°C. Interestingly, among three *B. weihenstephanensis* isolates included in the isolate set tested here, two did not show growth in SMB at 6°C over 21 days and one showed stochastic growth (three replicates, where two showed growth and one did not, averaged a 1.9 log increase between day 0 and day 21), despite the fact that *B. weihenstephanensis* is typically defined as a "cold growing" *Bacillus* spp. (Lechner et al., 1998). However, many previous studies only assessed growth in rich

media (e.g. BHI), which are likely to support better and more robust growth at low temperatures, as also shown here (discussed below). In addition, other studies (Guinebretière et al., 2008; Beno et al., unpublished data) also found that not all *B. weihenstephanensis* strains were able to grow at temperatures of 7°C or below. For example, Guinebretière et al. (2008) characterized 143 phylogenetic group VI isolates (*B. weihenstephanensis* and *B. mycoides*) and found 7 isolates unable to grow at 7°C according to a rapid screening test done on J-agar.

While the genera *Bacillus* and *Paenibacillus* represented the majority of isolates characterized here, the order Bacillales includes a number of genera, species, strains, and clonal groups that have been associated with dairy and dairy products and that show the ability to grow at low temperatures. Notably, the order Bacillales includes not only the families Bacillaceae (which includes the genera *Bacillus*, *Lysinibacillus*, *Oceanobacillus*, and *Psychrobacillus*), Paenibacillaceae (which includes the genus *Paenibacillus*), and Planococcaceae (which includes the genera *Solibacillus* and *Viridibacillus*), but also includes the family Listeriaceae, which includes the genus *Listeria*. *Listeria* is also well recognized for its ability to grow at low temperatures (Chan and Wiedmann, 2009; Schmid et al., 2009). Our data further showed that dairy associated Bacillales isolates in different families carry the capability to grow in SMB at 6°C. For example, we found that the one *V. arenosi* isolate evaluated (FSL H7-0596) grew at 6°C in SMB, consistent with Ivy et al. (2012) who also reported that isolates representing *V. arenosi* showed growth in SMB at 6°C (Ivy et al., 2012). Importantly, *Viridibacillus* isolates are rare among dairy Bacillales isolates (Ivy et al., 2012; Moreno Switt et al., 2014), but are described as able to grow at temperatures as low as 5°C (Albert et al., 2007). Other genera characterized in the present study included *Psychrobacillus* (2 isolates) as well as *Oceanobacillus*, *Soliobacillus*, and *Lysinibacillus* (one isolate for each). While these genera are rarely isolated from dairy associated sources, *Lysinibacillus*, *Oceanobacillus*, *and Solibacillus* spores have been isolated from bulk tank raw milk samples (Miller et al., 2015) and *Psychrobacillus*, *Oceanobacillus*, and *Lysinibacillus* spores have been isolated from raw milk and dairy powder samples (Kent et al., 2016). While none of

these genera were able to grow at 6°C in SMB, many of these genera have been reported as being capable of growing at low temperatures in media other than SMB (Albert et al., 2007; Krishnamurthi et al., 2010). For example, as indicated by its name, *Psychrobacillus* species often can grow at low temperatures, sometimes as low as -2°C (Krishnamurthi et al., 2010).

Our data also further support that ability to grow at low temperatures in SMB (as a model for fluid milk) varies across genera, species, and clonal groups in the order Bacillales. While in a number of cases cold growth (which we define here as ability to grow at temperatures below 10°C) is associated with specific taxa, in many cases isolates in the same taxon differ considerably in their ability to grow in SMB at 6°C. For example, as discussed above, one of the three *B. weihenstephanensis* isolates characterized here showed stochastic growth in SMB at 6°C, while the other two isolates did not show growth, consistent with a previous study by Guinebretière et al. (2008), who also showed the variable ability of *B. weihenstephanensis* to grow at 7°C on J-agar. Similarly, while many *Paenibacillus* isolates were able to grow in SMB at 6°C, only rarely were all isolates of a species able to grow in these conditions. Species with consistent ability to grow in SMB at 6°C, included the unnamed *Paenibacillus* sp. 8 and 14 where all 2 and 5 isolates in the respective species were able to grow under these conditions. Additionally, neither *P. glucanolyticus* isolate grew under these conditions. All other *Paenibacillus* spp. represented by two or more isolates (i.e., *P. odorifer* and 3 unnamed *Paenibacillus* spp.) included isolates that differed in their ability to grow in SMB at 6°C. For example, of the two isolates representing unnamed *Paenibacillus* sp. 6, one isolate (FSL H7-0443) showed growth in SMB at 6°C, while one (FSL R5-0923) was unable to grow under these conditions.

P. odorifer ***Represents a Very Common*** **Paenibacillus** ***Isolated from Fluid Milk and is Characterized with "Variable Growth" Patterns***

Based on initial classification using *rpoB* sequencing and AT classification, the plurality of *Paenibacillus* isolates in this study (30/58 *Paenibacillus* isolates) represented the species *P.*

odorifer. Refinement of species classification using ANIb analyses of WGS data (along with types strains) confirmed 27 out of the 30 isolates initially classified as *P. odorifer* as members of this species. The predominance of *P. odorifer* among dairy *Paenibacillus* isolates is consistent with data reported by Ivy et al. (2012), who evaluated 737 *Paenibacillus* isolates from dairy sources and found that 68.9% of these isolates were *P. odorifer*. Importantly, both the previous study by Ivy et al. (2012) and the present study found that *P. odorifer* showed considerable variation for ability to grow in SMB at 6°C. Among the 25 isolates that were confirmed as *P. odorifer* by WGS and that were characterized for growth in BHI and SMB at 6 and 10°C, all isolates showed growth in both media at 10°C and in BHI at 6°C, but only some isolates showed growth in SMB at 6°C. Similarly, there was no significant association between growth patterns in SMB at 6°C (as determined by QTC) and *P. odorifer* clades, which provided initial evidence that ability to grow in SMB at 6°C may not be linked to clear genetic markers. We also observed differences in growth patterns even among isolates with the same *rpoB* AT, which further suggests that ability of *P. odorifer* to grow in SMB at 6°C differs even among closely related isolates. In the present study, two isolates representing *rpoB* AT 46 (FSL H3-0464 and FSL J3-0153) were used for in-depth characterization of growth at low temperatures. Interestingly, FSL H3-0464 did not grow at 6°C in SMB whereas FSL J3-0153 grew 1.9 log after 21 days (relative to day 0). The other *rpoB* ATs that were included twice for this part of the study were consistent in growth capabilities; both pairs of isolates representing *rpoB* AT 27 and 35 grew at least 1 log in SMB at 6°C. Importantly, some isolates showed growth during the initial screen of 101 dairy-associated Bacillales, but not in the in depth characterization of *P. odorifer*. Some of this may be attributed to cultures presenting below the limit of detection. In these instances, we cannot determine the true extent of reduction, suggesting that some of these isolates may show even more die off than our numbers predict.

The observation that *P. odorifer* isolates differed in their ability to grow at 6°C in SMB, while all isolates were able to grow in BHI at 6°C and in SMB and BHI at 10°C indicates that growth media differentially affects that ability of *P. odorifer* to grow at 6°C and that some strains

are more resilient to media and temperature stress than others. While some *Paenibacillus* can grow at temperatures as low as 6°C, most *Paenibacillus* grow optimally at temperatures between 28 and 40°C (Vos et al., 2011). Low temperatures and media that lack necessary nutritional qualities have been documented to reduce growth capabilities of these bacteria. For example, Choma et al. (2000) reported that *B. cereus* isolates were unable to grow at lower temperatures in courgette broth, made from zucchini, which has less sugar and protein than J-broth, a rich medium that allowed for growth of these isolates (Choma et al., 2000). Alternatively, it is also possible that due to nutrient limitation in SMB at 6°C, *P. odorifer* are less likely to initiate growth, as may be predicted by the "scout model" of the bacterial life cycle (Epstein, 2009; Buerger et al., 2012).

While the initial observation that the ability of *P. odorifer* to grow at 6°C in SMB was not associated with specific clades already suggested that it may be challenging to identify genetic markers associated with the ability to grow in SMB at 6°C, we still performed formal analyses to identify association of genes, GO terms, and SNPs with ability of *P. odorifer* isolates to grow in SMB at 6°C; gene identification for these analyses was performed with both OrthoMCL as well as HMM, which was used to identify specific "psychrotolerance genes of interest. Not surprisingly, no genes, GO terms, or SNPs were found to be associated with an isolate's ability to grow in SMB at 6°C. However, the two *P. odorifer* clades identified here show distinct gene and GO term presence/absence patterns, which suggests adaptation of these clades to different environments. For example several clusters related to denitrification, as well as to molybdenum cofactor biosynthesis were overrepresented among clade A isolates. Interestingly, molybdenum and nitrate reductase enzymes have been associated with cold resistance in cereal crops (Sun et al., 2009). This may be relevant as all four isolates in QTC cluster 3 ("growth with short lag phase") grouped into clade A, which may suggest that clade A is more adapted to cold growth. Interestingly, isolates from clade A showed higher normalized counts after 14 day of growth in SMB at 6°C as compared to isolates in clade B (0.805 and -.789 log, respectively; $p = 0.0254$).

Paenibacillus ***spp. Associated with Fluid Milk Represents Considerable Species Diversity that Remain to be Formally Described***

While the genus *Paenibacillus* currently represents > 200 different species (LPSN, http://www.bacterio.net/paenibacillus.html), including a number of species that have been associated with dairy spoilage (Ivy et al., 2012; Moreno Switt et al., 2014; Grady et al., 2016), our study suggests that *Paenibacillus* isolates from dairy associated environments represent considerable genomic diversity and a number of species that have not yet been formally described. Specifically, based on ANIb analyses of the WGS data created here for 55 *Paenibacillus* isolates and 3 previously sequenced *Paenibacillus* isolates, these isolates represent 21 different species. Based on comparisons with type strains with available WGS data, only three of these 21 species could be identified as named species; in addition to *P. odorifer*, named species found among these isolates included *P. glucanolyticus* and *P. macerans*. While the remaining 18 species may include some species that have been previously described, but have no type strains with WGS data, it is likely that a number of these species may represent new species of *Paenibacillus*, found in dairy-associated sources, that have not yet been named. Future work to clarify the taxonomy of the *Paenibacillus* isolates in these unnamed species will require WGS of a number of type strains. This type of future work will not only advance our understanding of Paenibacillus taxonomy and diversity, but will also be essential to provide for better tools to detect and prevent fluid milk spoilage due to aerobic spore-formers.

CONCLUSIONS

Our study clearly shows the tremendous phenotypic and genomic diversity associated with aerobic spore-formers in the order Bacillales, which represent a key concern for spoilage of fluid milk. This diversity presents considerable challenges as we aim to control and reduce fluid milk spoilage. For example, our data suggest that a simple molecular assay that detects psychrotolerant spore-formers that may cause fluid spoilage will unlikely materialize due to the fact that organisms

with these types of phenotypic characteristics do not represent well defined clades or carry easily detectable “cold growth genes”. Similarly, the diversity of psychrotolerant Bacillales and their presence across different families, genera, and clades suggests that development of predictive model assessing fluid milk spoilage also requires in-depth data on these types of spoilage organisms. Future efforts to control fluid milk spoilage due to psychrotolerant spore-formers thus likely will require multiplexed sequence based detection and characterization approaches that both define subtypes and identify genes linked to relevant phenotypes (e.g., spoilage processes). The WGS data reported here not only provide an important starting point for future efforts to develop these types of tools, but also provide a valuable resource for additional *Paenibacillus* genome analysis efforts.

All Supplemental Tables relevant to this paper, as well as the original growth data for both the screen and in depth analyses of *P. odorifer* isolates, are available in Appendix B.

ACKNOWLEDGEMENTS: Sarah Beno was supported by a USDA National Needs Grant (2013-38420-20519). In addition, the New York State Dairy Promotion Board provided funding to support this research. The authors gratefully acknowledge the following members of the Cornell University Milk Quality Improvement Program: Maureen Gunderson for preparing the necessary media for this project and Rachel Evanowski, Rachel Miller, Shanna Smith, and Yumin Xu for their assistance with the collection of cold growth data.

REFERENCES

Aguilar, P. S., J. E. Cronan Jr., and D. de Mendoza. 1998. A *Bacillus subtilis* gene induced by cold shock encodes a membrane phospholipid desaturase. *J. Bacteriol*. 180(8):2194-2200.

Albert, R. A., J. Archambault, M. Lempa, B. Hurst, C. Richardson, S. Gruenloh, M. Duran, H. L. Worliczek, B. E. Huber, R. Rosselló-Mora, P. Schumann, and H.-J. Busse. 2007. Proposal of *Viridibacillus* gen. nov. and reclassification of *Bacillus arvi*, *Bacillus arenosi*, and *Bacillus neidei* as *Viridibacillus arvi* gen. nov., comb. nov., *Virdibacillus arenosi* comb. nov. and *Viridibacillus neidei* comb. nov. *Int. J. Sys. Evol. Microbiol*. 57:2729-2737. http://dx.doi.org/10.1099/ijs.0.65256-0.

Ash, C., F. G. Priest, and M. D. Collins. 1993. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*. *Antonie van Leeuwenhoek*. 64(3):253-260. http://dx.doi.org/10.1007/BF00873085.

Babraham Bioinformatics. FastQC: A quality tool for high throughput sequence data. Accessed 20 Jun 2017. http://bioinformatics.babraham.ac.uk/projects/fastqc.

Bairoch, A. and R. Apweiler. 2000. The SWISS-PROT protein sequence database and its

supplement TrEMBL in 2000. *Nucleic Acids Res*. 28(1):45-48. http://dx.doi.org/10.1093/nar/28.1.45

Bakersmans, C., P. W. Bergholz, D. F. Rodrigues, T. A. Vishnivetskaya, H. L. Ayala-del-Río, and J. M. Tiedje. 2012. Genomic and expression analyses of cold-adapted microorganisms. Pages 126-155 in Polar Microbiology: Life in a Deep Freeze. R. V. Miller and L. G. Whyte, ed. ASM Press, Washington, DC.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19(5):455-477. http://dx.doi.org/10.1089/cmb.2012.0021.

Barria, C., M. Malecki, and C. M. Arraino. 2013. Bacterial adaptation to cold. *Microbiology*. 159:2437-2443. http://dx.doi.org/10.1099/mic.0.052209-0.

Beno, S. M., M. J. Stasiewicz, A. D. Andrus, R. D. Ralyea, D. J. Kent, N. H. Martin, M. Wiedmann, and K. J. Boor. 2016. Development and validation of pathogen environmental monitoring programs for small cheese processing facilities. *J. Food Prot*. 79(12):2095-2106. http://dx.doi.org/10.4315/0362-028X.JFP-16-241.

Berge, O., M.-H. Guinebretière, W. Achouak, P. Normand, and T. Heulin. 2002. *Paenibacillus graminis* sp. nov. and *Paenibacillus odorifer* sp. nov., isolated from plant roots, soil and food. *Int. J. Sys. Evol. Micorbiol*. 52:607-616. http://dx.doi.org/10.1099/00207713-52-2-607.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114-2120. http://dx.doi.org/10.1093/bioinformatics/btu170.

Budde, I., L. Steil, C. Scharf, U. Volker, and E. Bremer. 2006. Adaptation of *Bacillus subtilis* to grow at low temperature: a combinded transcriptomic and proteomic appraisal.

Microbiology. 152:831-853. http://dx.doi.org/10.1099/mic.0.28530-0.

Buerger, S., A. Spoering, E. Gavrish, C. Leslin, L. Ling, and S. S. Epstein. 2012. Microbial scout hypothesis, stochastic exit from dormancy, and the nature of slow growers. *Appl. Environ. Microbiol.* 78(9):3221-3228. http://dx.doi.org/10.1128/AEM.07307.11. https://www.ers.usda.gov/webdocs/publications/43833/43680_eib121.pdf?v=41817.

Carlin, F., J. Brillard, V. Broussolle, T. Clavel, C. Duport, M. Jobin, M.-H. Guinebretière, S. Auger, A. Sorokine, and C. Nguyen-the. 2010. Adaptation of *Bacillus cereus*, an ubiquitous worldwide-distributed foodborne pathogen, to a changing environment. *Food Research Int.* 43(7):1885-1894. http://dx.doi.org/10.1016/j.foodres.2009.10.024.

Chan, J. Z.-M., M. R. Halachev, N. J. Loman, C. Constantinidou, and M. J. Pallen. 2012. Defining bacterial species in the genomic era: insights from the genus Acinetobacter. *BMC Microbiol*. 12:302. http://dx.doi.org/10.1186/1471-2180-12-302.

Chan, Y. C. and M. Wiedmann. 2009. Physiology and genetics of *Listeria monocytogenes* survival and growth at cold temperatures. *Crit. Rev. Food Sci. Nutr*. 49(3):237-253. http://dx.doi.org/10.1080/10408390701856272.

Chassaing, D. and F. Auvray. 2007. The lmo1078 gene encoding a putative UDP-glucose pyrophosphorylase is involve in growth of Listeria monocytogenes at low temperature. *FEMS Microbiol. Lett*. 275(1):31-37. http://dx.doi.org/10.1111/j.1574-6968.2007.00840.x

Choma, C., T. Clavel, H. Dominguez, N. Razafindramboa, H. Soumille, C. Nguyen-the, and P. Schmitt. 2000. Effect of temperature on growth characteristics of *Bacillus cereus* TZ415. *Int. J. Food Microbiol*. 55:73-77. http://dx.doi.org/10.1016/S0168-1605(00)00197-5.

Conesa, A. and S. Götz. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics*. http://dx.doi.org/10.1155/2008/619832.

Dsouza, M., M. W. Taylor, S. J. Turner, and J. Aislabie. 2014. Genome-based comparative analyses of Antarctic and temperate species of *Paenibacillus*. *PLoS ONE* 9(10):e108009. http://dx.doi.org/10.1371/journal.pone.0108009.

Eddy, S. R. 2011. Accelerated profile HMM searches. *PLoS Comp. Biol*. 7:e1002195. http://dx.doi.org/10.1371/journal.pcbi.1002195.

Epstein, S. S. 2009. Microbial awakenings: A theory of how microbes 'wake up' from dormancy could help to solve scientific mysteries and improve disease control. *Nature*. 457: 1083.

Finn, R. D., P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate, and A. Bateman. 2016. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* 44:D279-D285. http://dx.doi.org/10.1093/nar/gkv1344.

Francis, K. P., R. Mayr, F. Von Stetten, G. S. A. B. Stewart, and S. Scherer. 1998. Discrimination of psychrotrophic and mesophilic strains of the *Bacillus cereus* group by PCR targeting of major cold shock protein genes. *Appl. Envion. Microbiol*. 64(9):3525-3529.

Gardner, S. N., T. Slezak, and B. G. Hall. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics*. 31(17):2877-2878. http://dx.doi.org/10.1093/bioinformatics/btv271.

Grady, E. N., J. MacDonald, L. Liu, A. Richman, and Z. Yuan. 2016. Current knowledge and perspectives of *Paenibacillus*: a review. *Microb. Cell Fact*. 15:203 http://dx.doi.org/10.1186/s12934-016-0603-7.

Graumann, P., T. M. Wendrich, M. H. Weber, K. Schoröder, and M. A. Marahiel. 1997. A family of cold shock proteins in *Bacillus subtilis* is essential for cellular growth and for efficient protein synthesis at optimal and low temperatures. *Mol. Microbiol*. 25(4):742-756.

Guinebretière, M.-H., F. L. Thompson, A. Sorokin, P. Normand, P. Dawyndt, M. Ehling-Schulz, B. Svensson, V. Sanchis, C. Nguyen-the, M. Heyndrickx, and P. de Vos. 2008. Ecological diversification in the *Bacillus cereus* group. *Environ. Microbiol*. 10(4):851-865. http://dx.doi.org/10.1111/j.1462-2920.2007.01495.x.

Guinebretière, M.-H., O. Berge, P. Normand, C. Morris, F. Carlin, and C. Nguyen-the. 2001. Identification of bacteria in pasteurized zucchini purées stored at different temperatures and comparison with those found in other pasteurized vegetable purées. *Appl. Environ. Microbiol*. 67(10):4520-4530. http://dx.doi.org/10.1128/AEM.67.10.4520-4530.2001.

Gurevich, A., V. Saveliev, N. Vyahhi, and G. Tesler. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 29(8):1072-1075. http://dx.doi.org/10.1093/bioinformatics/btt086.

Hébraud, M. and P. Potier. 1999. Cold shock response and low temperature adaptation in psychrotrophic bacteria. *J. Mol. Microbiol. Biotechnol*. 1(2):211-219.

Henningsen, A. 2017. censReg: Censored Regression (Tobit) Models. R package version 0.5-26. https://CRAN.R-project.org/package=censReg.

Huck, J. R., N. H. Woodcock, R. D. Ralyea, and K. J. Boor. 2007. Molecular subtyping and characterization of psychrotolerant endospore-forming bacteria in two New York State fluid milk processing systems. *J. Food Prot*. 70(10):2354-2364.

Ivy, R. A., M. L. Ranieri, N. H. Martin, H. C. den Bakker, B. M. Xavier, M. Wiedmann, and K. J. Boor. 2012. Identification and characterization of psychrotolerant sporeformers associated with fluid milk production and processing. *Appl. Environ. Microbiol*. 78(6):1853-1864. http://dx.doi.org/10.1128/AEM.06535-11.

Kaan, T., G. Homuth, U. Mader, J. Bandow, and T. Schweder. 2002. Genome-wide transcriptional profiling of the *Bacillus subtilis* cold-shock response. *Microbiology*. 148:3441-3455. http://dx.doi.org/10.1099/00221287-148-11-3441.

Kent, D. J., K. Chauhan, K. J. Boor, M. Wiedmann, and N. H. Martin. 2016. Spore test parameters matter: Mesophilic and thermophilic spore counts detected in raw milk and dairy powders differ significantly by test method. *J. Dairy Sci*. 99(7):5180-5191. http://dx.doi.org/10.3168/jds.2015-10283.

Krishnamurthi, S., A. Ruckmani, R. Pukall, and T. Chakrabarti. 2010. *Psychrobacillus* gen. nov.

and proposal for reclassification of *Bacillus insolitus* Larkin & Stokes, 1967, *B. psychrotolerans* Abd-El Rahman et al., 2002 and *B. psychrodurans* Abd-El Rahman et al., 2002 as *Psychrobacillus insolitus* comb. nov., *Psychrobacillus psychrotolerans* comb. nov. and *Psychrobacillus psychrodurans* comb. nov. *Sys. Appl. Microbiol*. 33(7):367-373. http://dx.doi.org/10.1016/j.syapm.2010.06.003.

Lechner, S., R. Mayr, K. P. Francis, B. M. Pruβ, T. Kaplan, E. Wieβner-Gunkel, G. S. A. B. Stewart, and S. Scherer. 1998. *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. *Int. J. Sys. Bacteriol*. 48:1373-1382. http://dx.doi.org/10.1099/00207713-48-4-1373.

Leisch, F. 2006. A toolbox for K-Centroids cluster analysis. *Computational Statistics and Data Analysis*. 51(2):526-544. http://dx.doi.org/10.1016/j.csda.2005.10.006.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25(16):2078-2079. http://dx.doi.org/10.1093/bioinformatics/btp352.

Li, L., C. J. Stoeckert Jr., and D. S. Roos. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res*. 13:2178-2189. http://dx.doi.org/10.1101/gr.1224503.

List of prokaryotic names with standing in nomenclature (LPSN). 2017. Genus *Paenibacillus*. Accessed Jun 20, 2017. http://www.bacterio.net/paenibacillus.html.

Masiello, S. N., D. J. Kent, N. H. Martin, Y. H. Schukken, M. Wiedmann, and K. J. Boor. 2017. Longitudinal assessment of dairy farm management practices associated with the presence of psychrotolerant Bacillales spores in bulk tank milk on 10 New York State dairy farms. *J. Dairy Sci*. 100(11): 8783-8795. http://dx.doi.org/10.3168/jds.2017-13139.

Masiello, S. N., N. H. Martin, A. Trmčić, M. Wiedmann, and K. J. Boor. 2016. Identification and characterization of psychrotolerant bacteria isolated from pasteurized fluid milk. *J. Dairy Sci.* 99(1):130-140. http://dx.doi.org/10.3168/jds.2015-9728.

Masiello, S. N., N. H. Martin, R. D. Watters, D. M. Galton, Y. H. Schukken, M. Wiedmann, and K. J. Boor. 2014. Identification of dairy farm management practices associated with the presence of psychrotolerant sporeformers in bulk tank milk. *J. Dairy Sci*. 97(7):4083-4096. http://dx.doi.org/10.3168/jds.2014-7938.

Meer, R. R., J. Baker, F. W. Bodyfelt, and M. W. Griffiths. 1991. Psychrotrophic *Bacillus* spp. in fluid milk products: A review. *J. Food Prot*. 54:964-979. http://dx.doi.org/10.4315/0362-028X-54.12.969.

Miller, R. A., S. M. Beno, D. J. Kent, L. M. Carroll, N. H. Martin, K. J. Boor, and J. Kovac. 2016. *Bacillus wiedmannii* sp. nov., a psychrotolerant and cytotoxic *Bacillus cereus* group species isolated from dairy foods and dairy environments. *Int. J. Sys. Evol. Microbiol*. 66:4744-4753. http://dx.doi.org/10.1099/ijsem.0.001421.

Miller, R. A., D. J. Kent, K. J. Boor, N. H. Martin, and M. Wiedmann. 2015. Different management practices are associated with mesophilic and thermophilic spore levels in bulk tank raw milk. *J. Dairy Sci*. 98(7):4338-4351. http://dx.doi.org/10.3168/jds.2015-9406.

Moreno Switt, A. I., A. D. Andrus, M. L. Ranieri, R. H. Orsi, R. Ivy, H. C. den Bakker, N. H. Martin, M. Wiedmann, and K. J. Boor. 2014. Genomic comparison of sporeforming bacilli isolated from milk. *BMC Genomics*. 15:26. http://dx.doi.org/10.1186/1471-2164-15-26.

O'Leary, N. A., M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrun, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, and K. D. Pruitt. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic

expansion, and functional annotation. *Nucleic Acids Res*. 44:D733-D745. http://dx.doi.org/10.1093/narl/gkv1189.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ranieri, M. L., R. A. Ivy, R. Mitchell, E. Call, S. N. Masiello, M. Wiedmann, and K. J. Boor. 2012. Real-Time PCR detection of *Paenibacillus* spp. in raw milk to predict shelf life performance of pasteurized fluid milk products. *Appl. Environ. Microbiol*. 78(16):5855-5863. http://dx.doi.org/10.1128/AEM.01361-12.

Richter, M. and R. Rosselló-Móra. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA*. 106(45):19126-19131. http://dx.doi.org/10.1073/pnas.0906412106.

Rivas, R., C. Gutiérrez, A. Abril, P. F. Mateos, E. Martínez-Molina, A. Ventosa, and E. Velázquez. 2005. *Paenibacillus rhizospaerae* sp. nov., isolated from the rhizosphere of *Cicer arietinum*. *Int. J. Sys. Evol. Microbiol*. 55:1305-1309. http://dx.doi.org/10.1099/ijs.0.63513-0.

Scharl, T. and F. Leisch. 2006. The stochastic QT-clust algorithm: Evaluation of stability and variance on time-course microarray data. Pages 1015-1022. Compstat 2006- Proceedings in Computational Statistics. A. Rizzi and M. Vichi, ed., Physica Verlag, Heidelberg, Germany.

Schmid, B., J. Klumpp, E. Raimann, M. J. Loessner, R. Stephan, and T. Tasara. 2009. Role of cold shock proteins in growth of *Listeria monocytogenes* under cold and osmotic stress conditions. *Appl. Environ. Microbiol*. 75(6):1621-1627. http://dx.doi.org/10.1128/AEM.02154-08.

Shen, H. W., R. C. Yu, and C. C. Chou. 2007. Acid adaptation affects the viability of *Salmonella* Typhimurium during the lactic fermentation of skim milk and product storage. *Int. J. Food Microbiol.* 114(3):380-385. http://dx.doi.org/10.1016/j.ijfoodmicro.2006.09.033.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30(9):1312-1313. http://dx.doi.org/10.1093/bioinformatics/btu033.

Stenfors Arnesen, L. P., K. O'Sullivan, and P. E. Granum. 2007. Food poisoning potential of *Bacillus cereus* strains from Norweigian dairies. *Int. J. Food Microbiol*. 116:292-296. http://dx.doi.org/10.1016/j.ijfoodmicro.2006.12.021

Sun, X., C. Hu, Q. Tan, J. Liu, and H. Liu. 2009. Effects of molybdenum on expression of cold-responsive genes in abscisic acid (ABA)-dependent and ABA-independent pathways in winter wheat under low-temperature stress. *Ann Bot*. 104(2):345-356. http://dx.doi.org/10.1093/aob/mcp133.

Tatusova, T., M. DiCuccio, A. Badretdin, V. Chetvernin, E. P. Nawrocki, L. Zaslavsky, A. Lomsadze, K. D. Pruitt, M. Borodovsky, and J. Ostell. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res*. 44(14):6614-6624. http://dx.doi.org/10.1093/narlgkw569.

Trmčić, A., N. H. Martin, K. J. Boor, and M. Wiedmann. 2015. A standard bacterial isolate set for research on contemporary dairy spoilage. *J. Dairy Sci*. 98(8):5806-5817. http://dx.doi.org/10.3168/jds.2015-9490.

Ulusu, N. N. and E. F. Tezcan. 2001. Cold Shock Proteins. *Turk. J. Med. Sci*. 31:283-290.

Vos, P., G. Garrity, D. Jones, N. R. Krieg, W. Ludwig, F. A. Rainey, K.-H. Schleifer, and W. Whitman. 2011. Family Paenibacillaceae. Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes. Spinger Science and Business Media, Berlin, Germany.

Weber, M. H., W. Klein, L. Müller, U. M. Niess, and M. A. Marahiel. 2001. Role of the *Bacillus subtilis* fatty acid desaturase in membrane adaptation during cold shock. *Mol. Microbiol*. 39(5):1321-1329.

Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

Xia, B., H. Ke, and M. Inouye. 2001. Acquirement of cold sensitivity by quadruple deletion of

the *cspA* family and its suppression by PNPase S1 domain in *Escherichia coli*. *Mol. Microbiol.* 40(1):179-188.

Zhou, G., D. Zheng, L. Dou, Q. Cai, and Z. Yuan. 2010. Occurrence of psychrotolerant *Bacillus cereus* group strains in ice creams. *Int. J. Food Microbiol*. 137:143-145. http://dx.doi.org/10.1016/j.ijfoodmicro.2009.12.005.

CHAPTER 4

A NUMBER OF GENES ARE ASSOCIATED WITH PSYCHROTOLERANT BACILLUS CEREUS *GROUP ISOLATES*

In Preparation for Submission to Journal of Dairy Science

ABSTRACT

The *Bacillus cereus* group includes nine different species that include human pathogens as well as psychrotolerant strains that have been associated with fluid milk spoilage. In order to develop a better understanding of cold growth capabilities among dairy associated *B. cereus* groups isolates, we selected 22 genetically distinct representative isolates from a collection of 503 dairy associated isolates to characterize by whole genome sequencing and for their ability to grow at 6°C in Skim Milk Broth (SMB) and Brain Heart Infusion (BHI) broth. Growth experiments identified nine isolates that were able to grow in BHI (average of 1.45 and 2.07 log at days 14 and 21, respectively), including two isolates that were also able to grow in SMB. A Maximum Likelihood phylogeny built with genome wide core SNPs classified all isolates that showed ability to grow at 6°C in BHI into clade VI, which has previously been show to represent the species *B. weihenstephanensis* and *B. mycoides*. Analysis of correlations between gene presence/absence patterns (determined using OrthoMCL) and psychrotolerance identified 206 orthologous gene clusters that were overrepresented among isolates that were able to grow at 6°C in BHI ("cold growers"), including two clusters representing cold-shock proteins, which were found in 8/9 and 7/9 isolates that showed ability to grow at 6°C. Gene Ontology (GO) term analyses revealed 36 GO terms overrepresented in isolates able to grow at 6°C, including putrescine catabolic processes and putrescine transmembrane transporter activity. Hidden Markov Model searches finally identified three protein families previously linked to ability to grow at low temperature as significantly associated with the bacterial count on day 21 from cultures incubated at 6°C in BHI, including the Cold Shock Domain and Fatty Acid hydroxylases. Analyses of *cspA* sequences

revealed a positive association between cold growers and a previously identified psychrotolerant *cspA* signature sequence. Overall, our data not only support that *B. cereus* group clade IV, which includes all isolates classified as *B. weihenstephanensis* and *B. mycoides*, represents the primary clade of concern in refrigerated dairy products, but also identify a number of gene targets that could be used for specific detection or control of psychrotolerant *B. cereus* group detect spoilage organisms in the dairy value chain.

INTRODUCTION

The *Bacillus cereus* group includes nine closely related species (*B. anthracis*, *B. cereus*, *B. cytotoxicus*, *B. mycoides*, *B. pseudomycoides*, *B. thuringiensis*, *B. toyonensis*, *B. weihenstephanensis*, and *B. wiedmannii*) that can also be classified into seven phylogenetic clades with 4 clades representing a single species and 3 clades representing two or three different species (Guinebretière et al., 2008, Kovac et al., 2016). The *B. cereus* group species are associated with a variety of niches and are highly diverse with regard to cytotoxicity, presence of toxin genes, and growth temperature ranges (Guinebretière et al., 2008; EFSA BIOHAZ Panel, 2016). While some species are recognized as human pathogens (e.g., *B. anthracis*) (Didelot et al., 2009; Bottone, 2010), others (e.g., *B. thuringiensis*) are used as insecticides (Höfte and Whitely, 1989) or probiotics (e.g., *B. toyonensis*) (Jiménez et al., 2013; Kantas et al., 2015). Some species, in particular *B. weihenstephanensis*, are associated with food spoilage (Bartoszewicz et al., 2008; Hwang and Park, 2015) and are commonly isolated from fluid milk (Ivy et al., 2012; Saleh-Lakha et al., 2017). Species classification of *Bacillus cereus* group isolates often uses diagnostic species-specific phenotypic characteristics; for example, isolates with rhizoid morphology are classified as either *B. mycoides* or *B. pseudomycoides* (Nakamura and Jackson, 1995; Nakamura, 1998). While *B. cereus* species and strains are considered mesophilic (Koehler, 2009; Ceuppens et al., 2013; Jiménez et al., 2013), growth temperature ranges differ between species. For example, *B. cytotoxicus* is the only thermotolerant species in the group and isolates classified into this species are capable of growth at temperatures up to 50°C (Guinebretière et al., 2013). Importantly, growth capability at different temperatures has also been used for species classification, with *B. weihenstephanensis* defined by the ability of isolates to grow at 7 but not 43°C (Lechner et al., 1998). While *B. weihenstephanensis* was the first species in the *B. cereus* group observed to grow at refrigeration temperatures (Lechner et al., 1998), other *B. cereus* groups species have been reported to include isolates that can grow at temperatures below 7°C (Guinebretière et al., 2008; Soufiane and Côte, 2013; Miller et al., 2016). In addition, *B. wiedmannii* has recently been

described as a new *B. cereus* group species that is not only characterized by the ability to grow at refrigeration temperatures, but also its ability to produce toxins that are typically associated with *B. cereus* strains linked to foodborne illnesses (Miller et al., 2016).

Among the *B. cereus* group species, *B. weihenstephanensis* represent the species most often associated with fluid milk spoilage (Griffiths, 1992; Páčová et al., 2003; Bartoszewicz et al., 2008). *B. weihenstephanensis*, as well as potentially other cold growing *Bacillus* spp., thus represent a particular concern for the dairy industry, as do *Paenibacillus* species that can grow at refrigeration temperatures (Huck et al., 2008; Ivy et al., 2012; Gopal et al., 2015). It has been proposed that *B. weihenstephanensis* can be identified based on signature single nucleotide polymorphisms (SNPs) at nucleotides 4 and 9 of *cspA*, which encodes a cold shock protein (Francis et al., 1998). However, Soufiane and Côte (2013) found that these signature SNPs were not unique to *B. weihenstephanensis*, but were also present in isolates classified as *B. mycoides* and that some isolates classified as *B. cereus* also harbored this signature sequence. This group (Soufiane and Côte, 2013) additionally reported that several housekeeping genes (e.g., *gmk*, *glpF*, and *tpi*) had DNA signatures unique to psychrotolerant *B. cereus* group strains. Psychrotolerant strains were defined by their ability to show growth at 7°C in LB broth and included strains identified as *B. weihenstephanensis*, *B. cereus*, and *B. mycoides*. As such, we hypothesize that *cspA* is not the only gene responsible for the ability of a *B. cereus* group isolate to grow at low temperatures. This hypothesis is consistent with previous research that identified proteins that were associated with, or responsible for, ability to grow at refrigeration temperatures in other members of the order Bacillales. For example, DEAD box helicases, chaperone DnaJ, and the low temperature requirement protein A (LtrA) were found to be associated with the ability of *Paenibacillus* isolates to grow at 6°C in SMB (Moreno Switt et al., 2014). In addition, a number of proteins have been associated with the ability of *Listeria monocytogenes* to grow at low temperatures, including cold-shock proteins (CSPs), DEAD box helicases, and several enzymes modifying lipid membrane fluidity (Chan and Wiedmann, 2009).

Based on the importance of psychrotolerant *B. cereus* group species as fluid milk spoilage organisms, and possible concerns about psychrotolerant *B. cereus* group species and strains that could cause foodborne illness when present in dairy products, further studies on the cold growth capabilities of *B. cereus* group isolates are needed in order to improve our ability to control psychrotolerant members of this group in dairy products. The objectives of this study were (i) to quantify growth of *B. cereus* group isolates at refrigeration temperature and (ii) to identify genes and genetic markers associated with the ability of an isolate to grow at refrigeration temperatures. The data from this study will provide for improved classification of isolates with regard to ability to grow at refrigeration temperatures and also will provide essential information that can be used to develop improved DNA-based assays for rapid and specific identification of psychrotolerant *B. cereus* group isolates throughout the dairy supply chain.

MATERIALS AND METHODS

Isolate Selection

Isolates for this study were selected to focus on members of the *B. cereus* group that have the ability to grow at refrigeration temperatures. Isolate selection was based on an initial cold growth screen of isolates representing all 42 *rpoB* allelic types found among 503 dairy associated *B. cereus* group isolates in the Food Safety Lab collection (http://www.foodmicrobetracker.com). The initial screen was performed by spread plating 100 µl of overnight cultures representing the 42 selected isolates on Brain Heart Infusion (**BHI**) agar; presence of colonies after incubation at 6°C for 21 days in both biological replicates was considered indicative of ability to grow at 6°C. The screen identified 14 isolates (each representing a different *rpoB* AT) that screened positive for ability to grow at 6°C; all of these isolates (which represented clades II and VI) were included in the study reported here. To allow for comparative studies, we additionally included nine isolates representing *B. cereus* group subtypes and clades that do not have the ability to grow at 6°C. These additional isolates included (i) one isolate that grouped into clade II, but did not show

ability to grow at 6°C and (ii) representatives of *B. cereus* group clades I (1 isolate), III (3 isolates), and IV (4 isolates, including the *B. cereus s.s.* type strain, ATCC 14579 (Ivanova et al., 2003)); all of these clades only include isolates that did not show ability to grow at 6°C. No isolates from clades V or VII (representing *B. toyonensis* and *B. cytotoxicus*, respectively) were included as clade V isolates were rare in our collection of 503 *B. cereus* group dairy isolates, while clade VII isolates were not represented at all in this collection.

Quantification of Growth at 6°C in Brain Heart Infusion Broth or Skim Milk Broth

The 23 selected *B. cereus* group isolates were tested for their ability to grow at 6°C in the nutrient-rich BHI broth or a medium representative of fluid milk, Skim Milk Broth (**SMB**). For these growth experiments, isolates were streaked, in triplicate, onto BHI agar from frozen glycerol stocks, followed by incubation for 18-24 h at 32°C. One colony of each replicate was inoculated into 5 ml of BHI broth and incubated for 24 h at 32°C (without aeration). These cultures were used to inoculate pre-cooled tubes that contained 5 ml of either BHI or SMB with a starting inoculum of approximately 10^2 CFU/ml. Tubes were incubated at 6°C, representing slight temperature abuse, for 21 days (without aeration). On days 0, 14, and 21, cultures were spiral plated in duplicate using an Autoplate 5000 (Advanced Instruments, Inc., Norwood, MA) onto Standard Plate Count (**SPC**) agar. After spiral plating, SPC agar plates were incubated at 32°C for 24 ± 2 h, followed by enumeration of colonies using a Q-Count Colony Counter (Advanced Instruments, Inc.). Samples without any bacterial growth on either of the duplicate plates, indicating bacterial counts below the detection limit of 10 CFU/ml, were arbitrarily and conservatively counted as 1 CFU/ml.

Quality Threshold Clustering

Bacterial counts for days 14 and 21 were normalized to day 0 and then averaged for each isolate. Using the qtclust function (Scharl and Leisch, 2006) of the flexclust package (Leisch,

2006) in R, data was clustered according to Quality Threshold Clustering (**QTC**). Data was clustered using a radius of 1.8 log CFU/ml for 14 and 21 day counts in both BHI and SMB.

DNA Extraction and Preparation for Whole Genome Sequencing

Among the 23 isolates included in this study, whole genome sequence (**WGS**) data had already been generated for 8 isolates, including (i) FSL J3-0113 (Miller et al., 2016), and (ii) 7 isolates for which WGS data were reported by Kovac et al. (2016). For the remaining 15 isolates, whole genome sequencing was performed as part of the study reported here. For these isolates, DNA was extracted from cultures grown in BHI media using the QIAamp DNA Mini kit (Qiagen, Valencia, CA) according to a modified protocol that included a 45 min lysis step with 180 µl of 20 mg/ml lysozyme in a 37°C water bath. DNA was eluted in 50 µl of Tris-HCl (pH 8.0). The concentration of double-stranded DNA was normalized to a concentration of 1 ng/µl and submitted to the Cornell University Institute of Biotechnology Genomics Facility (Ithaca, NY) for Nextera XT DNA library preparation. Samples were sequenced in two different Illumina HiSeq runs with 2 x 100 bp paired-end sequencing; these two runs were targeted to yield 83x and 113x coverage, respectively.

Read Processing, Quality Control, Genome Assembly, and Annotation

Low quality bases, reads, and Nextera XT adapters were trimmed using the default settings of Trimmomatic v0.33 (Bolger et al., 2014). We assessed short read quality using FastQC (v0.11.2) (Babraham Bioinformatics). Using SPAdes v3.6.2 and a variety of k-mer sizes (21, 33, 55, 77, 99), genomes were assembled *de novo* (Bankevich et al., 2012). QUAST was used to verify the quality of the assembled draft genomes (Gurevich et al., 2013). Using BBMap v35.49 and computing the average depth using SAMtools, average coverage was determined by mapping the reads against draft genomes (Li et al., 2009). Sequence reads and assembled draft genomes were submitted to SRA and NCBI's WGS database, respectively (see Table 4.1 for details), using

the prokaryotic genome annotation pipeline (Tatusova, 2016).

TABLE 4.1: Twenty-three *B. cereus* group isolates representing 22 different *rpoB* ATs were selected for cold growth quantification over 21 days at 6°C in SMB and BHI broth.

Isolate	Species Identification	WGS Accession	Clade[a]	*rpoB* AT
FSL H8-0534	*B. pseudomycoides*	MUAQ00000000	I	148
FSL W8-0169	*B. wiedmannii*	LOBC00000000[c]	II	61
FSL K6-0069	*B. wiedmannii*	LOBB00000000[c]	II	194
FSL M8-0091	*B. wiedmannii*	MUAM00000000	II	410
FSL J3-0113	*B. wiedmannii*	LXFN00000000[d]	II	417
FSL M8-0117	*B. cereus*	LONG00000000[c]	III	308
FSL W8-0483	*B. cereus*	LOMU00000000[c]	III	120
FSL W8-0050	*B. cereus*	LOMR00000000[c]	III	125
FSL M8-0473	*B. cereus* ATCC 14579^T	MUAP00000000	IV	158
FSL R5-0811	*B. cereus*	MUAO00000000	IV	158
FSL W8-0268	*B. cereus*	LOMS00000000[c]	IV	92
FSL K6-1030	*B. cereus*	MUAU00000000	IV	556
FSL M7-0669	*B. weihenstephanensis*	MUAK00000000	VI	3
FSL H7-0683	*B. mycoides* [b]	MUAR00000000	VI	75
FSL H7-0926	*B. weihenstephanensis*	LOBD00000000[c]	VI	90
FSL M7-1219	*B. weihenstephanensis*	MUAL00000000	VI	97
FSL H8-0485	*B. weihenstephanensis*	MUAJ00000000	VI	132
FSL H8-0492	*B. weihenstephanensis*	MUAS00000000	VI	134
FSL R5-0708	*B. weihenstephanensis*	MUAN00000000	VI	257
FSL M7-0109	*B. weihenstephanensis*	MUAH00000000	VI	273

FSL J3-0123	*B. weihenstephanensis*	MUAG00000000	VI	513
FSL E2-0214	*B. weihenstephanensis*	MUAT00000000	VI	531
FSL W7-1108	*B. mycoides* [b]	MUAI00000000	VI	342

[a] Phylogenetic clades were assigned according to WGS clustering (Guinebretière et al., 2008; Kovac et al., 2016)

[b] *B. mycoides* was assigned to isolates in phylogenetic clade VI with rhizoid morphology

[c] Sequenced by Kovac et al. (2016)

[d] Sequenced by Miller et al. (2016)

SNP Detection and Phylogeny Construction

SNPs were called using kSNP3 (Gardner et al., 2015). The k-mer size of 31 was selected using Kchooser. A maximum likelihood (**ML**) tree was constructed using RAxML v.8.0 (Stamatakis, 2014) and the core SNPs detected by kSNP3. The ML tree was constructed using a time-reversible (GTR) model with gamma-distributed sites (GAMMA) and 1000 bootstrap repetitions. The phylogenetic tree was edited using FigTree v.1.4.2.

OrthoMCL and Gene Ontology (GO) Term Annotation

All 23 *B. cereus* group genomes were analyzed using OrthoMCL (Li et al., 2003) with an inflation value of 2.5 to find ortholog clusters (groups of orthologous genes found across multiple isolates). A representative protein sequence for each of the 9,885 identified clusters was used for gene ontology (**GO**) annotation using Blast2GO (Conesa and Götz, 2008) searches against two databases. First, all protein sequences were searched against the SWISS-PROT database (Bairoch and Apweiler, 2000). Then, protein sequences with no GO terms mapped using SWISS-PROT were searched against the RefSeq database (O'Leary et al., 2016). The outputs from SWISS-PROT and RefSeq were combined and the assigned GO terms were linked to their respective ortholog clusters and to each member of the ortholog cluster. OrthoMCL clusters that were overrepresented among cold growers were assigned gene names by running BLAST using the protein sequences against RefSeq and SWISS-PROT.

Gene Presence/Absence Analysis and Gene Enrichment

Using the ortholog clusters from the OrthoMCL output, counts of genomes where each gene was present or absent were computed for isolates classified as showing "growth" or "reduction" in BHI; "growth" was defined as at least a 1 log increase in bacterial numbers at either day 14 or 21, while "reduction" was defined as at least a 1 log decrease in bacterial numbers at either day 14 or 21. Presence/absence data for each OrthoMCL cluster (gene) were used to

generate 2x2 tables, which were analyzed using two-sided Fisher's exact tests. Odds ratios were computed and p-values were adjusted using the False Discovery Rate (**FDR**). To identify GO terms that were over-or under-represented among genomes of isolates that were classified as showing "growth" or "reduction" in BHI, the number of genes classified as a given GO term were summed in each genome and then across all the genomes for isolates classified as showing "growth" or "reduction". This approach was used to generate 2x2 tables for each GO term, which were used to run two-sided Fisher's exact tests and compute odds ratios as described above. FDR < 0.05 were considered statistically significant.

Cold Shock Protein Sequence Analysis and Classification

Amino acid sequences of putative Csp proteins identified by SWISS-PROT and Refseq were compared against previously published sequences that are unique to each Csp protein, namely CspA, CspB, CspC, CspD, CspE (Mayr el al., 1996; Francis et al., 1998; and Schindler et al., 1999) and an uncharacterized Csp protein described in Mayr et al. In addition, the mesophilic and psychrotolerant variants of CspA were also characterized. According to Francis et al. (1998), a CspA protein with the sequence MAVTGQVKWFNNEKGFGF is found among mesophilic *B. cereus* group isolates, while psychrotolerant isolates display the CspA sequence, MTVTGQVKWFNNEKGFGFIEVPG. Using Fisher's exact tests, we tested for association between phenotypic psychrotolerance and presence of the psychrotolerant *cspA* sequence. P-values < 0.05 were considered statistically significant.

Identification of Proteins Related to Psychrotolerance

Twelve Hidden Markov Model (**HMM**) protein domains previously shown to be associated with psychrotolerance and ability to grow at low temperatures were obtained from Pfam 26.0 protein families' database (see Supplemental Table 4.1; Finn et al., 2016). HMMER v. 3.2 (Eddy, 2015) was used to search these HMM models against the genomes for the 23 isolates

characterized here. Linear regression was run in R to determine whether protein family matches per genome are associated with relative growth at 6°C in BHI on day 21 (defined as bacterial numbers at day 21 – bacterial numbers at day 0).

RESULTS

Among 23 B. cereus *Group Isolates, 9 Grew at 6°C in BHI Broth, but Only 2 Grew at 6°C in SMB*

Bacterial enumeration data were used to calculate bacterial growth at 6°C after 14 and 21 days, which was defined as bacterial numbers at days 14 and 21 relative to bacterial numbers at day 0; all growth data represent the average of three replicates. Growth patterns were categorized numerically into the following three groups (i) isolates with at least 1 log growth at either day 14 or 21 ("growth"); (ii) isolates with at least 1 log reduction at either day 14 or 21 and no more than 1 log growth at either time point ("reduction"); (iii) isolates that did not classify into either (i) or (ii) ("no growth"). In BHI, 9 of the 23 isolates showed growth (average of 1.45 and 2.07 log at days 14 and 21, respectively), while in SMB only two isolates showed growth (average of 1.09 and 1.61 log at days 14 and 21, respectively) (see Table 4.2 for details). Both isolates that showed growth in SMB also showed growth in BHI. All nine isolates that showed growth in BHI represented WGS phylogenetic clade VI, which contains *B. weihenstephanensis* and *B. mycoides* isolates; the other two isolates that grouped into clade VI showed "reduction" in BHI (average of 1.94 and 1.04 log reduction at days 14 and 21, respectively).

In BHI, all 14 isolates that did not show "growth", fell into the category "reduction" with an average of 2.45 and 2.50 log reductions at days 14 and 21, respectively. In SMB, the isolates that did not fall into the category "growth" represented the category "reduction" (17 isolates, average of 2.14 and 2.16 log reduction at days 14 and 21, respectively) and the category "no growth" (4 isolates). Interestingly, 13/14 isolates that showed "reduction" in BHI also showed "reduction" in SMB (see Table 4.2).

Quality Threshold Clustering Reveals 3 Cold Growth Patterns for **B. cereus** *Group Isolates*

To further identify groups of isolates that shared similar growth patterns in SMB and BHI at 6°C, we performed QTC, which revealed 3 clusters into which 21/23 isolates were grouped. QTC Cluster 1 contains 12 isolates that predominantly were numerically categorized as "reduction" in both BHI and SMB; only one isolate in QTC Cluster 1 did not show "reduction". Specifically, FSL H7-0926 showed "reduction" in BHI and showed 0.77 and 0.44 log lower bacterial counts at days 14 and 21, respectively, which classified it into the "no growth" group (Table 4.2, Figures 4.1 and 4.2A). On average, QTC Cluster 1 isolates showed 2.59 and 2.70 log reduction at day 14 and 21, respectively, in BHI and 2.25 and 1.95 log reduction at day 14 and 21, respectively, in SMB.

Five isolates were classified into QTC Cluster 2 (see Table 4.2, Figures 4.1 and 4.2B for details); all isolates in this cluster showed growth of at least 1 log in BHI (mean growth of 1.29 and 2.19 log for day 14 and 21, respectively). Two of the isolates in QTC Cluster 2 grew at least 1 log in SMB and the remaining three were categorized as "no growth" in SMB. QTC Cluster 3 contains four isolates that were able to grow in BHI (mean growth of 1.66 and 1.91 log for day 14 and 21, respectively), but were classified as showing "reduction" in SMB (see Table 4.2, Figures 4.1 and 4.2C for details). While all nine isolates in QTC Clusters 2 and 3 represented clade VI, the two additional clade VI isolates tested here fell into QTC Cluster 1 or did not cluster (see Table 4.2).

The two isolates that were not classified into any of the three QTC clusters (FSL H7-0683 and FSL W8-0169, clades VI and II, respectively) could be manually added to QTC Cluster 1 using numerical categorization, as each isolate showed "reduction" in both BHI and SMB (see Table 4.2 and Figure 4.1 for details). These isolates appear to not have been classified into QTC Cluster 1 as the magnitude of reduction under some conditions (SMB or BHI) or at a given time point was lower than that associated with other isolates in QTC Cluster 1; for example, FSL H7-0683 only showed 0.07 log reduction at day 21 in BHI, while the QTC Cluster 1 isolate with the

least reduction of bacterial numbers at day 21 still showed 1.71 log reduction for that time point.

TABLE 4.2. Isolates were categorized in their ability to grow at 6°C in both BHI and SMB using numerical categorization and Quality Threshold Clustering

	BHI			SMB			
Isolate	Day 14	Day 21	Growth Category [a]	Day 14	Day 21	Growth Category [a]	QTC Cluster
FSL H7-0926	-2.45	-2.01	Die off	-0.77	-0.44	No growth	1
FSL H8-0534	-1.37	-1.81	Die off	-2.41	-1.53	Die off	1
FSL J3-0113	-2.89	-3.49	Die off	-1.94	-2.83	Die off	1
FSL K6-0069	-2.50	-2.70	Die off	-2.68	-1.06	Die off	1
FSL K6-1030	-2.39	-2.39	Die off	-2.06	-2.51	Die off	1
FSL M8-0091	-2.97	-2.97	Die off	-2.77	-2.77	Die off	1
FSL M8-0117	-2.15	-1.71	Die off	-2.21	-1.67	Die off	1
FSL M8-0473	-2.09	-2.53	Die off	-2.86	-1.53	Die off	1
FSL R5-0811	-3.40	-3.40	Die off	-2.43	-2.14	Die off	1
FSL W8-0050	-2.92	-2.92	Die off	-1.83	-2.27	Die off	1
FSL W8-0268	-3.46	-3.46	Die off	-3.03	-2.59	Die off	1
FSL W8-0483	-2.51	-3.06	Die off	-2.05	-2.05	Die off	1
FSL J3-0123	0.74	1.85	Growth	-0.72	-0.98	No growth	2
FSL M7-0109	2.08	2.44	Growth	0.01	0.62	No growth	2
FSL M7-0669	1.43	2.65	Growth	1.05	1.88	Growth	2
FSL M7-1219	1.64	2.16	Growth	1.13	1.34	Growth	2
FSL R5-0708	0.55	1.87	Growth	0.11	0.43	No growth	2
FSL E2-0214	1.82	2.40	Growth	-2.24	-2.79	Die off	3
FSL H8-0485	1.18	1.78	Growth	-2.64	-3.09	Die off	3
FSL H8-0492	1.43	2.05	Growth	-1.25	-1.52	Die off	3
FSL W7-1108	2.22	1.41	Growth	-1.81	-1.71	Die off	3

FSL H7-0683	-1.42	-0.07	Die off	-2.15	-2.79	Die off	Did not cluster
FSL W8-0169	-1.79	-2.52	Die off	-0.07	-1.84	Die off	Did not cluster

[a] Categories defined according to: (i) isolates that showed at least 1 log CFU/ml above base line at either day 14 or 21 ("growth"); (ii) isolates that showed at least 1 log CFU/ml decline at either day 14 or 21 and did not show more than 1 log growth at either time point ("die off"); (iii) No more than 1 log CFU/ml growth or die off at either time point and not classified into either (i) or (ii) ("no growth"). Importantly, in instances where the bacterial count was below the limit of detection (i.e., < 10 CFU/ml), we arbitrarily and conservatively set this reduction at zero log (i.e., 1 CFU/ml).

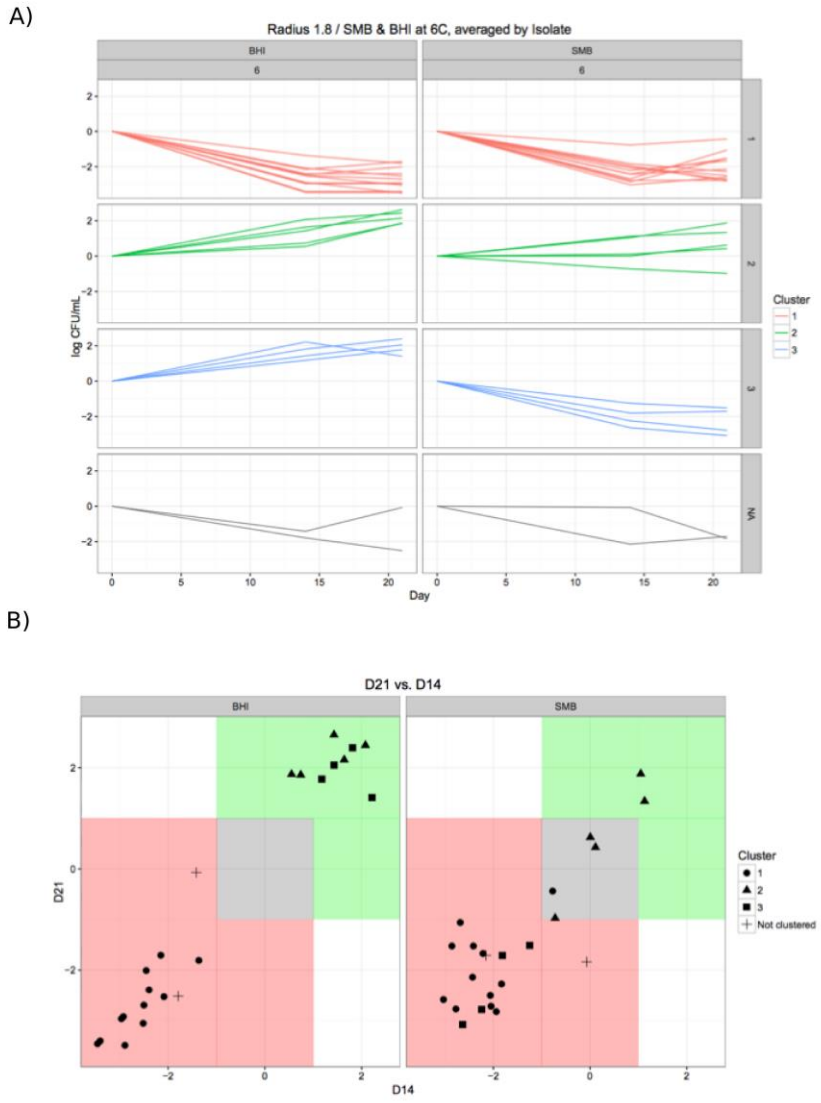


FIGURE 4.1: Growth patterns in SMB and BHI broth at 6°C for 23 *B. cereus* group isolates. (A) Isolates were grouped according to growth in both SMB and BHI broth using Quality Threshold Clustering with a radius of 1.8. Growth data of three replicates for each isolate were averaged for each isolate and clustered according to their growth in these conditions. Cluster groups are color coded, and correspond with growth patterns. Cluster 1, marked in red, is comprised of 12 isolates that generally showed die off in both BHI and SMB at 6°C. Isolates (n=5) in cluster 2, marked in green, generally showed at least 1 log higher bacterial counts at either day 14 or 21 as compared to

day 0 in both BHI and SMB. Cluster 3, marked in blue, is comprised of 4 isolates that showed growth in BHI but die off in SMB at 6°C. There were also two isolates, in gray, that did not group using Quality Threshold Clustering. (B) Growth data (relative to day 0) at each time point was averaged for each of the 23 isolates (3 replicates) and plotted for day 14 on the x-axis and for day 21 on the y-axis. Both BHI and SMB are displayed. Different shapes (e.g., circle, triangle) on the graph indicate each QTC group. The area in green represents growth, while the area in red represents die off. The area shaded in gray represents no growth. BHI is on the left and SMB is on the right. Importantly, in instances where the bacterial count was below the limit of detection (i.e., < 10 CFU/ml), we arbitrarily and conservatively set this reduction at zero log (i.e., 1 CFU/ml).

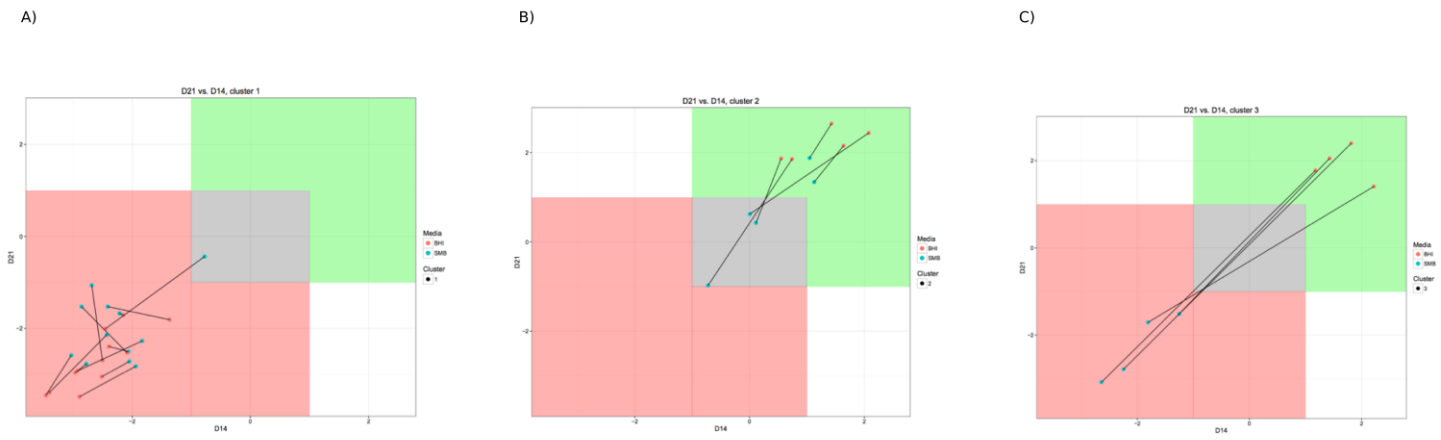


FIGURE 4.2: Growth data at each time point was plotted for each QTC cluster, individually. (A) Cluster 1 growth data are plotted for both BHI and SMB. Lines connect data from the same isolates, with the red dot indicating the day 14 x day 21 bacterial counts in BHI, with the blue dot indicating the day 14 x day 21 bacterial counts in SMB. As in Figure 4.1B, the green shaded area is indicative of growth, while gray represents no growth and red indicates die off. Consistent with Figure 4.1A, isolates from QTC cluster 1 generally died off in both BHI and SMB. (B) Cluster 2 growth data are plotted for both BHI and SMB. As in Figure 4.2A, lines connect data from the same isolates, with the red dot indicating bacterial counts in BHI and the blue dot indicating bacterial counts in SMB. Consistent with Figure 4.1A, isolates from QTC cluster 2 showed growth in BHI and did not die off in SMB (2 isolates showed growth while 3 showed no growth). In all cases, counts in BHI were higher than counts in SMB. (C) Cluster 3 growth data are plotted for both BHI and SMB. As in Figure 4.2A, lines connect data from the same isolates, with the red dot

indicating bacterial counts in BHI and the blue dots representing bacterial counts in SMB. Consistent with Figure 4.1A, isolates from QTC cluster 3 showed growth in BHI but die off in SMB. Importantly, in instances where the bacterial count was below the limit of detection (i.e., < 10 CFU/ml), we arbitrarily and conservatively set this reduction at zero log (i.e., 1 CFL/ml).

Only Isolates from Phylogenetic Clade VI are able to Grow at 6°C in either SMB or BHI Broth

Numerical growth classification and QTC clustering both support that phylogenetic clade VI, which includes the species *B. weihenstephanensis* and *B. mycoides* (Figure 4.3), includes all isolates that are able to grow at 6°C. Based on numerical categorization, the 11 isolates in clade VI included (i) 2 isolates that showed "growth" for both BHI and SMB, (ii) 7 isolates that showed "growth" for BHI and either "reduction" or "no growth" in SMB (4 and 3 isolates, respectively), and (iii) two isolates that showed "reduction" in BHI and either "no growth" or "reduction" in SMB. Importantly, all *B. cereus* group isolates that showed either "growth" or "no growth" classified into clade VI. All isolates in the other clades tested (i.e., clades I, II, III, and IV) consistently showed reduction in both SMB (mean decrease of 2.20 and 2.07 log at day 14 and 21, respectively) and BHI at 6°C (mean decrease of 2.54 and 2.75 log at day 14 and 21, respectively).

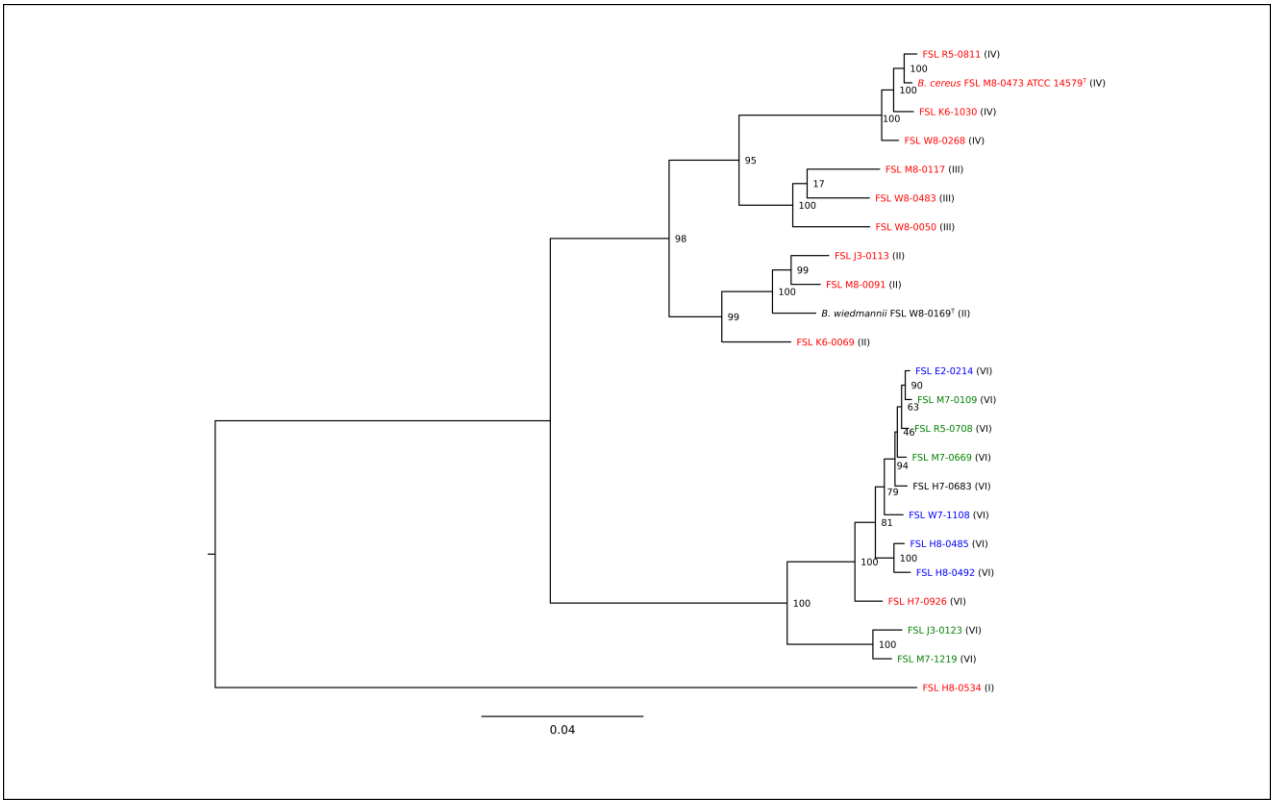


FIGURE 4.3: Phylogenetic tree constructed from the core SNPs identified in the genomes of 23 *B. cereus* isolates. The maximum likelihood tree was constructed using a general time-reversible model with gamma-distributed sites and 1,000 bootstrap repetitions. The isolates group into five previously described phylogenetic clades (Guinebretière et al., 2008; Kovac et al., 2016). QTC clusters representing isolates with similar growth patterns in BHI and SMB at 6°C (see Figure 4.2A) are mapped onto the phylogenetic tree, using the same colors for each cluster as shown in Figure 4.2A. Isolates marked in red represent those that showed die off in both BHI and SMB at 6°C. Isolates marked in green generally showed growth in both BHI and SMB at 6°C while those marked in blue showed growth in BHI but not in SMB at 6°C. Importantly, these groups were set with the adjustment where, in instances where the bacterial count was below the limit of detection (i.e., < 10 CFU/ml), we arbitrarily and conservatively counted this as 1 CFU/ml.

Genes Encoding for Invasion Proteins and Cold-Shock Proteins, and protein families involved in putrescine metabolism are Overrepresented among* B. cereus *Group Isolates Showing Growth at 6°C in BHI

Overall, OrthoMCL identified 9,885 different orthologous clusters (which can essentially be thought of as genes) among the 23 genomes analyzed. Fisher's exact tests with an FDR correction was used to identify genes that were overrepresented among isolates able to grow in BHI at 6°C; we did not test for genes overrepresented among isolates that showed growth at 6°C in SMB, as only two isolates fell into this category. These analyses revealed 206 orthologous gene clusters significantly overrepresented in isolates able to grow at 6°C in BHI and 182 orthologous gene clusters significantly overrepresented in isolates unable to grow in these conditions (Supplemental Tables 4.2 and 4.3). Importantly, we did not identify any orthologous gene clusters that were found in all 9 isolates that were able to grow in BHI at 6°C and were absent from all 14 isolates that did not grow in BHI at 6°C. However we identified 87 clusters (including some discussed in more detail below) that were found among all 9 "cold growers" and only 2/14 "non cold growers"; all of these genes clusters were specific for clade VI and were only found among the 11 clade VI strains characterized here.

Interestingly, two orthologous clusters representing cold-shock proteins (cluster_4979 and cluster_5279, see Supplemental Table 4.2) were overrepresented in isolates able to grow at 6°C (Supplemental Table 4.2). These clusters were identified as genes encoding YdjO (98% coverage to *B. subtilis* YdjO) and CspC (100% coverage to *B. subtilis* CspC), respectively. *ydjO* (cluster_4979) was identified in 8/9 isolates able to grow at 6°C (all except for FSL W7-1108), while *cspC* (cluster_5279) was identified in 7/9 isolates that could grow under these conditions (all except for FSL H8-0492 and FSL M7-0669); importantly though, as the 23 genomes are not closed, lack of identification of a gene in a given isolate may not always reflect true absence, but could represent a lack of appropriate sequence coverage. Interestingly, *ydjO* was found in one isolate unable to grow at 6°C (FSL H7-0926) and *cspC* was identified in two isolates unable to

grow at 6°C (FSL H7-0926 and FSL H7-0683). Both of these genes were only identified in isolates from phylogenetic clade VI, which includes *B. weihenstephanensis* and *B. mycoides* isolates, regardless of ability to grow at 6°C; specifically, *ydjO* and *cspC* were found in 10 and 8 of the 11 clade VI isolates, respectively.

Other potentially interesting orthologous clusters overrepresented in the 9 isolates able to grow at 6°C in BHI include (i) two genes encoding invasion proteins (cluster_4691 and 4799), (ii) two genes annotated as encoding homologues of the damage inducible protein DinB (cluster_4457 and 4953), (iii) two genes encoding glyoxalases (cluster_4700 and 4937), and (iv) two genes encoding GNAT family acetyltransferases (cluster_4699 and 4966) (Supplemental Table 4.2). Both orthologous clusters encoding for invasion proteins were found in all 9 cold growing isolates and in the two non-cold growing isolates (FSL H7-0926 and FSL H7-0683) that are also part of phylogenetic clade VI; no specific pre-existing gene names could be assigned to these two clusters. Orthologous clusters 4457 and 4953, which were annotated as the damage-inducible protein DinB, showed 98 and 99% coverage to *dinB* in several species in the *B. cereus* group, respectively. Coverage measures percentage of amino acids in a protein sequence that match amino acids in a reference sequence. Coverage however, is not the same as similarity or identity. Therefore, all genes mentioned here are not confirmed, but are hypothesized based on coverage results from BLAST searches. Cluster_4457 was identified in all cold growers, as well as in three isolates that did not show growth in 6°C in BHI broth; these three isolates however showed growth at 6°C on BHI agar (Miller et al., unpublished data). This cluster was identified in all 11 clade VI isolates and in isolate FSL J3-0113 (clade II). Cluster_4953 was identified in 8/9 isolates that showed growth at 6°C (all except for FSL H8-0482) and in 2/14 isolates that did not show growth at 6°C (FSL H7-0926 and FSL H7-0683). Overall, cluster_4953 was identified in 10/11 isolates classified into phylogenetic clade VI. Among the two orthologous clusters annotated as glyoxalases, no specific gene name could be assigned to cluster_4937, but cluster_4700 was identified as MqoO (98% coverage in *B. subtilis* with equal coverage matches to paralogs MhqA

and MhqE). Cluster_4937 was found in 8/9 isolates able to grow at 6°C in BHI (all expect for FSL W7-1108), while cluster_4700 was found in all isolates able to grow at 6°C in BHI. Cluster_4937 and Cluster_4700 were found in 10/11 and 11/11 clade VI isolates, respectively. Among the two clusters identified as GNAT family acetyltransferases, cluster_4699 showed 97% coverage to YoaA of *B. subtilis*, while cluster_4966 could not be assigned a specific gene name (no coverage match > 95%). *yoaA* was present in all 9 cold-growing isolates and in 2 non cold-growing isolates, while Cluster_4966 was identified in 8 of the 9 cold-growing isolates and in 2 non cold-growing isolates. *yoaA* was found in all 11 clade VI isolates while cluster_4966 was identified in 10 of the 11 clade VI isolates (excluding FSL M7-1219); neither of these clusters were identified in isolates outside of clade VI.

GO term analyses revealed 36 terms that were significantly overrepresented in isolates that were able to grow at 6°C and 14 terms that were significantly overrepresented in isolates unable to grow at 6°C in BHI broth (see Supplemental Tables 4.4 and 4.5 for details). GO terms found to be overrepresented in the genomes of cold-growing *B. cereus* group isolates included the terms "putrescine catabolic process" (GO:0009447) and "putrescine transmembrane transporter activity" (GO:0015489) (see Supplemental Table 4.4 for complete list). GO terms found to be overrepresented in the genomes of *B. cereus* group isolates unable to grow at 6°C in BHI broth include "alkanesulfonate transporter activity" (GO:0042959), "alkanesulfonate monooxygenase activity" (GO:0008726), and "alkanesulfonate catabolic process" (GO:0046306) (see Supplemental Table 4.5 for complete list).

Targeted searches for protein families and SNPs previously associated with cold growth capabilities identify additional genetic markers for cold growth capabilities in* B. cereus *group isolates.

To further test for specific genetic markers associated with cold growth, we performed HMM searches for 12 protein families that had previously been identified as being associated with

cold growth and cold adaptation. Members of 11 of these 12 protein families were identified in at least one of the 23 *B. cereus* group genomes tested (see Supplemental Table 4.6 for details). For eight protein families, members were identified in all 23 genomes, although the number of hits in a given genome sometimes differed considerably. For example, either 6 or 7 hits for the Cold Shock Domain (CSD) were identified among each of the 23 different genomes; isolates in QTC Cluster 2, which showed growth in BHI and either growth or no growth in SMB, averaged 6.8 hits whereas all other isolates averaged 6.3 hits. The three HMM families for which we found hits in less than 23 genomes included LtrA, Caps_synth-CapC, and FA_desaturase_2. LtrA was identified in only two genomes (FSL M8-0473 [ATCC 14579^{T}] and FSL R5-0811), both representing *rpoB* AT 158. Caps_synth-CapC and FA_desaturase_2 were found at least once in 22 and 20 of the 23 genomes, respectively (see Supplemental Table 4.6 for details).

For three HMM families, we found a significant positive association between number of hits and log growth at day 21 in BHI; these families include CSD (p=0.01318), FA_hydroxylase (p=0.01094), and YdjO (p=0.001657). Isolates able to grow at 6°C in BHI averaged 6.8, 2.1 and 2.7 hits for CSD, FA_hydroxylase, and YdjO members, while isolates unable to grow under these conditions averaged 6.2, 1.5, and 1.9 hits, respectively. Furthermore, isolates from QTC Cluster 2 averaged more hits of these protein families than isolates in other QTC clusters. Surprisingly, the HMM family DEAD showed a significant negative association (p=0.02547) with growth in BHI at 6°C; isolates classified as showing "reduction" averaged more hits for DEAD (37.5) as compared to isolates that were classified as showing "growth" (35.9).

Analysis of the *cspA* sequences from each of the *B. cereus* group isolates reviewed here revealed a significant association (p=0.00337) between isolates with the psychrotolerant phenotype and those with the previously defined psychrotolerant cspA signature (Francis et al., 1998) by Fisher's exact test. The psychrotolerant *cspA* sequence was found in 12 isolates, including all 11 clade VI isolates and the one clade I isolate tested. The mesophilic *cspA* sequence was identified in the remaining *B. cereus* group genomes.

DISCUSSION

The *B. cereus* group includes a number of organisms relevant to dairy foods, including foodborne pathogens (for example, *B. cereus* strains expressing emetic or diarrheal toxins) and spoilage organisms of particular importance for fluid milk (e.g., *B. weihenstephanensis*). While a number of recent studies (Kovac et al., 2016; Warda et al., 2016; Zhang et al., 2017) have combined phenotypic and genomics-based approaches to allow for improved characterization and identification of *B. cereus* group clonal groups, strains, and isolates that have the ability to cause human disease, limited data are available on the genomic basis of cold growth capabilities among *B. cereus* group isolates. An improved understanding of cold growth capabilities and their genomic basis in the *B. cereus* group is needed for the dairy industry in order to allow for (i) improved risk assessments, (ii) improved detection and characterization methods that can rapidly detect and identify *B. cereus* group strains that can grow at refrigeration temperatures, and (iii) development of improved control strategies specifically for refrigerated dairy products that may permit *B. cereus* group growth (e.g. fluid milk, certain types of high pH cheeses). Our study provides important data characterizing cold growth capabilities among *B. cereus* group isolates specifically obtained from dairy associated sources. Our data specifically indicate that (i) ability to grow at 6°C in liquid media over typical HTST fluid milk shelf life seems to be limited to strains classified into *B. cereus* group clade VI, which represents the species *B. weihenstephanensis* and *B. mycoides* and that (ii) certain genomic features and genes are linked to ability of *B. cereus* group isolates to grow at 6°C, which will facilitate development of improved detection methods for these organisms.

Ability to Grow at 6°C in SMB or Rich Media is Limited to Some, but not All, Isolates in Phylogenetic Clade VI, which Represents B. weihenstephanensis *and* B. mycoides

While our initial screen for growth on BHI agar identified 14 *B. cereus* group isolates able to grow at 6°C (11 and 3 representing clades VI and II, respectively), only 9 and 2 of these isolates

also showed growth at 6°C in BHI broth and SMB, respectively. Interestingly, while three of the four isolates in phylogenetic clade II, which represents *B. wiedmannii*, showed ability to grow at 6°C on BHI agar, which is consistent with previous studies (Miller et al., 2016; Miller et al., unpublished data), none of these isolates grew at 6°C in BHI broth or SMB. *B. wiedmannii* was first described in 2016 as a psychrotolerant member of the *B. cereus* group (Miller et al., 2016) where 5 of 11 clade II isolates showed growth after 21 days of incubation at 5°C on BHI agar and all strains showed growth at 10°C. Combined, these data suggest a need for further characterization of *B. wiedmannii* for its ability to grow in different solid and liquid dairy products (e.g., fluid milk, fresh cheeses) under refrigeration conditions. This is particularly important since recent studies (Kovac et al., 2016; Miller et al., 2016) have shown that *B. wiedmannii* isolates carry virulence toxin genes associated with diarrheal disease (*hblACD* and *nheABC*) and show cytotoxicity in a HeLa cell culture model (Miller et al., 2016). Additionally, *B. wiedmannii* isolates have been shown to produce both HBL and NHE toxins (Kovac et al., 2016; Miller et al., 2016) using the Duopath Cereus Enterotoxins immunological lateral flow assay (Merck Millipore). The potential virulence of *B. wiedmannii* should be further assessed in dairy relevant conditions.

While nine isolates, which all were classified into *Bacillus cereus* group clade VI, showed growth in BHI broth at 6°C, only 2 of these isolates also showed the ability to grow in SMB at 6°C. This is consistent with previous experiments by our group, which found that among 28 *Paenibacillus* spp. that were able to grow in BHI at 6°C, only 16 were able to grow in SMB at this temperature (Beno et al., unpublished data). Future experiments may be needed to characterize and compare the ability of psychrotolerant Bacillales to grow in SMB and fluid milk in order to better assess food safety and spoilage risk associated with the presence of different Bacillales (e.g., *Paenibacillus*, *B. wiedmannii*, *B. weihenstephanensis*) in fluid milk. This is particularly important as autoclaving of SMB, as performed here, may result in Maillard reactions, which have been linked to antimicrobial activity (Ledl and Schleicher, 1990; Hauser et al., 2014). Clade VI, which

included all nine isolates that showed growth at 6°C in BHI broth, has previously been shown to represent the species *B. weihenstephanensis* and *B. mycoides* (Guinebretière et al., 2008). While a number of *B. cereus* group species have been reported to include isolates that grow at low temperatures (Stenfors and Granum 2001; Soufiane and Côte, 2013; Miller et al., 2016), our data are consistent with previous reports that *B. weihenstephanensis* isolates typically are able to grow at refrigeration temperatures (Lechner et al., 1998) and observations that isolates classified into this species have repeatedly been linked to spoilage issues in fluid milk (Páčová et al., 2003; Ivy et al., 2012). Likewise, our data show ability of *B. mycoides* to grow at 6ºC. *B. mycoides* was previously identified as capable of growth at 7ºC (Guinebretière et al., 2008), but is not classically defined as psychrotolerant; meanwhile, *B. weihenstephanensis*, which was discovered more than a century after *B. mycoides* was described in 1886 (Lewis, 1932), has been well described as psychrotolerant (Lechner et al., 1998; Stenfors and Granum, 2001). Among all 11 clade VI isolates tested here, only 2 isolates were not able to grow at 6ºC. Interestingly, previous studies have also found that not all *B. weihenstephanensis* show growth at low temperature. For example, a 2008 study by Guinebretière et al. reported 7/143 *B. weihenstephanensis* isolates were unable to grow at 7°C on J-agar, a rich medium.

Interestingly, all isolates outside of clade VI showed reductions in vegetative cell numbers in both BHI and SMB after incubation for 14 to 21 days at 6ºC, as did some clade VI isolates, particularly when incubated in SMB. While these findings may suggest die-off, reduction of vegetative cell numbers, as determined by plating on SPC agar, could also be caused by *B. cereus* group cells entering spore form during exposure to cold stress. Although sporulation due to cold stress has not been studied in *Bacillus* spp., a characterization study of *Clostridium thermocellum* JW20 found that vegetative cells entered spore form when the temperature fell below 45°C over several hours (Freier et al., 1988). Interestingly, a more recent study of *C. thermocellum* found that another strain (ATTC 27405) does not enter spore form at decreased temperatures (Mearls et al., 2012). Future experiments will thus be needed to assess sporulation of different Bacillales species

and isolates under cold stress exposure at dairy relevant conditions as well as the signals required for these cells to re-enter the vegetative status.

Overrepresentation of Selected Genes Linked to Cold Growth in Phylogenetic Clade VI Isolates Identifies Potential Targets for Detection of Isolates More Likely to Grow at Low Temperatures

Overall, 206 genes identified using OrthoMCL and 3 domains identified by HMM had a significant positive association with *B. cereus* group isolates that could grow at 6°C in BHI. Similar work done in *Paenibacillus* (Beno et al., unpublished data) did not find any associations between SNPs, genes, or GO terms and the ability of an isolate to grow at 6°C in SMB. Because we were unable to identify specific genes associated with the ability of an isolate to grow at 6°C in SMB, it is difficult to develop detection methods for *Paenibacillus* (notably, *P. odorifer*) that cause dairy spoilage. However, the *B. cereus* group has a number of genes and GO terms significantly associated with an isolate's ability to grow in BHI at 6°C. Importantly, all *Paenibacillus* isolates were able to grow in BHI at 6°C whereas *B. cereus* group isolates were not necessarily able to grow at 6°C, even in a rich medium.

In the present study, *cspA* sequences were examined to look for SNPs common to psychrotolerant strains in the *B. cereus* group. Although a significant association was found between cold growers and the psychrotolerant *cspA* signature sequence, it important to note that all isolates that grew were from phylogenetic clade VI (representing *B. mycoides* and *B. weihenstephanensis*) but that the psychrotolerant *cspA* signature was found in a clade I *B. pseudomycoides* isolate (FSL H8-0534), as well. Never has a study found a *B. pseudomycoides* isolate that was able to grow at low temperatures. Of similar interest, none of the clade II isolates (representing *B. wiedmannii*) harbored the psychrotolerant *cspA* sequence. Since *B. wiedmannii* is defined as a psychrotolerant species (Miller et al., 2016) it may be necessary to find a different gene to target when screening for psychrotolerant *B. cereus* group isolates.

Aside from *cspA*, there are numerous genes associated with cold-growing *B. cereus* group

isolates. HMM analyses revealed 4 protein families significantly associated with cold-growers, including the Cold Shock Domain (CSD), DEAD box RNA helicases, Fatty Acid hydroxylases, and the cold-inducible protein, YdjO. The cold shock domain is commonly associated with isolates that can grow at low temperatures, especially *B. weihenstephanensis* (Lechner et al., 1998; Francis et al., 1998). Fatty acid hydroxylases have also been previously linked to cold growth. The cell membrane fatty acid profiles change at low temperature to adjust membrane fluidity (Chan and Wiedmann, 2009; Mansilla et al., 2014; Diomandé et al., 2015). The cold-inducible protein YdjO is documented as cold-shock induced, but is functionally uncharacterized (Kaan et al., 2002). More "hits" of these protein families (CSD, fatty acid hydroxylases, and YdjO) were found in isolates that showed growth in BHI at 6°C. However, DEAD box RNA helicases, which have been previously reported to be associated with cold growth (Chan and Wiedmann, 2009; Barria et al., 2013; Moreno Switt et al., 2014), had more "hits" in isolates unable to grow in BHI at 6°C. Typically, these helicases help facilitate translation during exposure to low temperature (Chan and Wiedmann, 2009). While DEAD box helicases have often been associated with cold growth, these helicases are also associated with other stress modifications, as noted in a 2015 review, including reduced iron and oxidative stress (Redder et al.)

OrthoMCL analyses revealed cold-shock proteins were overrepresented in cold-growing *B. cereus* group isolates. Cold-shock proteins have been studied in the *B. cereus* group (Brillard et al., 2010) and a signature sequence of one particular cold-shock gene, *cspA*, has been used to identify *B. weihenstephanensis* (Lechner et al., 1998; Francis et al., 1998; Stenfors and Granum, 2001). Several GO terms were also significantly overrepresented in cold-growing *B. cereus* group isolates. These include aldehyde dehydrogenase and putrescine processes (e.g., putrescine transmembrane transporter activity, putrescine catabolic process). Interestingly, while the presence of putrescine, a polyamine, has been associated with ability of *Arabidopsis* and other plants to acclimate to cold temperatures (Cuevas et al., 2008; Palma et al., 2014; Lou et al., 2016), catabolism of spermidine, another polyamine, has been shown to be necessary for *E. coli* growth at

low temperature (Limsuwun and Jones, 2000). It has been suggested that, at low temperatures, excess spermidine replaces the Mg^{+2} bound to ribosomes resulting in their inactivation (He et al., 1993; Limsuwun and Jones, 2000) and therefore polyamine degradation at low temperatures may be needed to alleviate this toxic effect.

CONCLUSIONS

Isolates in the *B. cereus* group have different abilities to grow at 6°C. While previous studies have suggested that psychrotolerant isolates exist across a variety of species (Stenfors and Granum, 2001; Soufiane at Côte, 2013; Miller et al., 2016), the present study found only isolates from phylogenetic clade VI were able to grow at 6°C in either BHI or SMB. Genetic analyses revealed a number of genes and GO terms significantly associated with cold growing isolates from this study and confirmed that the psychrotolerant *cspA* allele is indeed associated with cold growing *B. cereus* group isolates. While the presence of the psychrotolerant DNA signature sequence of *cspA* was positively associated with cold growing *B. cereus* group isolates, it is important to consider that only clade VI isolates showed growth here. The psychrotolerant *cspA* allele captured all clade VI isolates (even those that did not grow in this study) and the clade I *B. pseudomycoides* isolate, FSL H8-0534, which also did not grow at 6°C in either BHI or SMB. Meanwhile, *B. wiedmannii*, a previously described psychrotolerant species with pathogenic potential is not detected here. For fluid milk, screening for the psychrotolerant specific *cspA* allele may be sufficient, but this may not be true for all dairy products. Future research focused on cold growing capabilities of *B. cereus* group isolates in solid media (such as cheese) will be valuable to the dairy industry to properly assess risk in different dairy products.

All Supplemental Tables relevant to this paper, as well as the original growth data and growth screen results for *B. cereus* group isolates, are available in Appendix C.

ACKNOWLEDGEMENTS: The New York State Dairy Promotion Board, through the New York State Department of Agriculture and Markets (Albany, NY), provided funding for this project. Sarah Beno was supported by a USDA National Needs Grant (2013-38420-20519). The authors gratefully acknowledge Maureen Gunderson, Cornell University Milk Quality Improvement Program, for preparing all necessary media for these experiments.

REFERENCES

Babraham Bioinformatics. FastQC: A quality tool for high throughput sequence data. http://bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 12 Apr 2017.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol*. 19(5):455-477. http://dx.doi.org/10.1089/cmb.2012.0021.

Bariroch, A. and R. Apweiler. 2000.The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research*. 28(1):45-48.

Barria, C., M. Malecki, and C. M. Arraino. 2013. Bacterial adaptation to cold. *Microbiology*. 159:2437-2443. http://dx.doi.org/10.1099/mic.0.052209-0.

Bartoszewicz, M., B. M. Hansen, and I. Swiecicka. 2008. The members of the *Bacillus cereus* group are commonly present contaminants of fresh and heat-treated milk. *Food Microbiol*. 25: 588-596. http://dx.doi.org/10.1016/j.fm.2008.02.001.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114-2120.

http://dx.doi.org/10.1093/bioinformatics/btu170.

Bottone, E. J. 2010. *Bacillus cereus*, a volatile human pathogen. *Clin. Microbiol. Rev.* 23(2):382-398. http://dx.doi.org/10.1128/CMR.00073-09.

Brillard, J., I. Jéhanno, C. Dargaignaratz, I. Barbosa, C. Ginies, F. Carlin, S. Fedhila, C. Nguyen-The, V. Broussolle, and V. Sanchis. 2010. Identification of *Bacillus cereus* genes specifically expressed during growth at low temperatures. *Appl. Envrion. Microbiol*. 76(8):2562-2573. http://dx.doi.org/10.1128/AEM.02348-09.

Ceuppens, S., N. Boon, and M. Uyttendaele. 2013. Diversity of *Bacillus cereus* group strains is reflected in their broad range of pathogenicity and diverse ecological lifestyles. *FEMS Microbiol. Ecol.* 84:433-450. http://dx.doi.org/10.1111/1574-6941.12110.

Chan, Y. C. and M. Wiedmann. 2009. Physiology and genetics of *Listeria monocytogenes* survival and growth at low temperatures. 2009. *Crit. Rev. Food Sci. Nutr*. 49(3):237-253. http://dx.doi.org/10.1080/10408390701856272.

Conesa, A. and S. Götz. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics*. http://dx.doi.org/10.1155/2008/619832.

Cuevas, J. C., R. López-Corbollo, R. Alcázar, X. Zarza, C. Koncz, T. Altabella, J. Salinas, A. F. Tiburcio, and A. Ferrando. 2008. Putrescine is involved in *Arabidopsis* freezing tolerance and cold acclimation by regulating abscisic acid levels in response to low temperature. *Plant Physiol*. 148(2):1094-1105. http://dx.doi.org/10.1104/pp.108.122945.

Didelot, X., M. Barker, D. Falush, and F. G. Priest. 2009. Evolution of pathogenicity in the *Bacillus cereus* group. *Sys. Appl. Microbiol*. 32:81-90. http://dx.doi.org/10.1016/j.syapm.2009.01.001.

Diomandé, S. E., M.-H. Guinebretière, B. De Sarrau, C. Nguyen-The, V. Broussolle, and J. Brillard. 2015. Fatty acid profiles and desaturase-encoding genes are different in thermo- and psychrtotolerant strains of the *Bacillus cereus* group. *BMC Res. Notes*. 8:329. http://dx.doi.org/10.1186/s1304-015-1288-4.

Eddy, S. R. 2015. Accelerated profile HMM searches. PLoS Comp. Biol. 7:e1002195. http://dx.doi.org/10.1371/journal.pcbi.1002195.

EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards). 2016. Scientific opinion on the risks for public health related to the presence of *Bacillus cereus* and other *Bacillus* spp. including *Bacillus thuringiensis* in foodstuffs. *EFSA Journal*. 14(7):4524. http://dx.doi.org/10.2903/j.efsa.2016.4524.

Finn, R. D., P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Slazar, J. Tate, and A. Bateman. 2016. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* 44:D279-D285. http://dx.doi.org/10.1093/nar/gkv1344.

Francis, K. P., R. Mayr, F. Von Stetten, G. S. A. B. Stewart, and S. Scherer. 1998. Discrimination of psychrotrophic and mesophilic strains of the *Bacillus cereus* group by PCR targeting of major cold shock protein genes. *Appl. Envion. Microbiol*. 64(9):3525-3529.

Freier, D., C. P. Mothershed, and J. Wiegel. 1988. Characterization of *Clostridium thermocellum* JW20. *Appl. Environ. Microbiol*. 54(1):204-211.

Gardner, S. N., T. Slezak, and B. G. Hall. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics*. 31(17):2877-2878. http://dx.doi.org/10.1093/bioinformatics/btv271.

Gopal, N., C. Hill, P. R. Ross, T. P. Beresford, M. A. Fenelon, and P. D. Cotter. 2015. The prevalence and control of *Bacillus* and related spore-forming bacteria in the dairy industry. *Front. Microbiol*. 6:1418. http://dx.doi.org/10.3389/fmicb.2015.01418.

Griffiths, M. W. 1992. *Bacillus cereus* in liquid milk and other milk products. *Bull. Int. Dairy Fed.* 36-39.

Guinebretière, M.-H., S. Auger, N. Galleron, M. Contzen, B. De Sarrau, M.-L. De Buyser, G. Lamberet, A. Fagerlund, P. E. Granum, D. Lereclus, P. De Vos, C. Nguyen-the, and A.

Sorokin. 2013. *Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* group occasionally associated with food poisoning. *Int. J. Sys. Evol. Microbiol*. 63: 31-40. http://dx.doi.org/10.1099/ijs.0.030627-0.

Guinebretière, M.-H., F. L. Thompson, A. Sorokin, P. Normand, P. Dawyndt, M. Ehling-Schultz, B. Svensson, V. Sanchis, C. Nguyen-The, M. Heyndrickx, and P. De Vos. 2008. Ecological diversification in the *Bacillus cereus* group. *Environ. Microbiol*. 10(4):851-865. http://dx.doi.org/10.1111/j.1462-2920.2007.01495.x.

Gurevich, A., V. Saveliev, N. Vyahhi, and G. Tesler. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 29(8):1072-1075. http://dx.doi.org/10.1093/bioinformatics/btt086.

Hauser, C., U. Müller, T. Sauer, K. Augner, and M. Pischetsrieder. 2014. Maillard reaction products as antimicrobial components for packaging films. *Food Chem*. 145:608-613. http://dx.doi.org/10.1016/j.foodchem.2013.08.083

He, Y., K. Kashiwagi, J. Fukuchi, K. Terao, A. Shirahata, and K. Igarashi. 1993. Correlation between the inhibition of cell growth by accumulated polyamines and the decrease of magnesium and ATP. *Eur J. Biochem*. 217(1):89-96.

Höfte, H. and H. R. Whitely. 1989. Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Mol. Biol. Rev*. 53(2):242-255.

Huck, J. R., M. Sonnen., K. J. Boor. 2008. Tracking heat-resistant, cold-triving fluid milk spoilage bacteria from farm to packaged product. *J. Dairy Sci*. 91(3):1218-1228. http://dx.doi.org/10.3168/jds.2007-0697.

Hwang, J.-Y. and J.-H. Park. 2015. Characteristics of enterotoxin distribution, hemolysis, l ecithinase, and starch hydrolysis of *Bacillus cereus* isolated from infant formulas and ready-to-eat foods. *J. Dairy Sci*. 98: 1652-1660. http://dx.doi.org/10.3168/jds.2014-9042.

Ivanova, N., A. Sorokin, I. Anderson, N. Galleron, B. Candelon, V. Kapatral, A. Bhattacharyya, G. Reznik, N. Mikhailova, A. Lapidus, L. Chu, M. Mazur, E. Goltsman, N. Larsen, M.

D'Souza, T. Walunas, Y. Grechkin, G. Pusch, R. Haselkorn, M. Fonstein, S. D. Ehrlich, R. Overbeek, and N. Kyrpides. 2003. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature*. 423:87-91. http://dx.doi.org/10.1038/nature01582.

Ivy, R. A., M. L. Ranieri, N. H. Martin, H. C. den Bakker, B. M. Xavier, M. Wiedmann, and K. J. Boor. 2012. Identification and characterization of psychrotolerant sporeformers associated with fluid milk production and processing. *Appl. Envrion. Microbiol*. 78(6):1853-1864. http://dx.doi.org/10.1128/AEM.06536-11.

Jiménez, G., M. Urdiain, A. Cifuentes, A. López-López, A. R. Blanch, J. Tamames, P. Kämpfer, A.-B. Kolstø, D. Ramón, J. F. Martínez, F. M. Codoñer, and R. Rosselló-Móra. 2013. Description of *Bacillus toyonensis* sp. nov., a novel species of the *Bacillus cereus* group, and pairwise genome comparisons of the species of the group by means of ANI calculations. *Int. J. Sys. Appl. Microbiol*. 36: 383-391. http://dx.doi.org/10.1016/j.syampm.2013.04.008.

Kaan, T., G. Homuth, U. Mäder, J. Bandow, and T. Schweder. 2002. Genome-wide transcriptional profiling of the *Bacillus subtilis* cold-shock response. *Microbiology*. 148:3441-3455.

Kantas, D., V. G. Papatsiros, P. D. Tassis, I. Giavasis, P. Bouki, and E. D. Tzika. 2015. A feed additive containing *Bacillus toyonensis* (Toyocerin(®)) protects against enteric pathogens in postweaning piglets. *J. Appl. Microbiol*. 118(3):727-738. http://dx.doi.org/10.1111/jam.12729.

Koehler, T. M. 2009. *Bacillus anthracis* physiology and genetics. *Mol. Aspects Med*. 30(6):386-396. http://dx.doi.org/10.1016/j.mam.2009.07.004.

Kovac, J., R. A. Miller, L. M. Carroll, D. J. Kent, J. Jian, S. M. Beno, and M. Wiedmann. 2016. Production of hemolysin BL by *Bacillus cereus* group isolates of dairy origin is associated with whole-genome phylogenetic clade. *BMC Genomics*. 17:581-597.

Lechner, S., R. Mayr, K. P. Francis, B. M. Prüβ, T. Kaplan, E. Wieβner-Gunkel, G. S. A. B. Stewart, and S. Scherer. 1998. *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. *Int. J. Sys. Bacteriol*. 48:1373-1382.

Ledl, F. and E. Schleicher. 1990. New aspects of Maillard reactions in foods and in the human body. *Angewandte Chemie – International Edition.* 29:565-594. http://dx.doi.org/10.1002/anie.199005653.

Leisch, F. 2006. A toolbox for K-Centroids cluster analysis. *Computational Statistics and Data Analysis.* 51(2): 526-544.

Lewis, I. M. 1932. Dissociation and life cycle of *Bacillus mycoides*. *J. Bacteriol*. 24(5):381-421.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25(16):2078-2079. http://dx.doi.org/10.1093/bioinformatics/btp352.

Li, L., C. J. Stoeckert Jr., and D. S. Roos. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*. 13:2178-2189. http://dx.doi.org/10.1101/gr.1224503.

Limsuwun, K. and P. G. Jones. 2000. Spermidine acetyltransferases is required to prevent spermidine toxicity at low temperatures in *Escherichia coli*. *J. Bacteriol*. 182(19):5373-5380.

Lou, Y.-R., M. Bor, J. Yan, A. S. Preuss, and G. Jander. 2016. Arabidopsis NATA1 acetylates putrescine and decreases defense-related hydrogen peroxide accumulation. *Plant Physiol*. 171:1443-1455. http://dx.doi.org/10.1104/pp.16.00446.

Mansilla, M. C., L. E. Cybulski, D. Albanesi, and D. de Mendoza. 2004. Control of membrane lipid fluidity by molecular thermosensors. *J. Bacteriol*. 186(20):6681-6688. http://dx.doi.org/10.1128/JB.186.20.6681-6688.2004.

Mearls, E. B., J. A. Izquierdo, and L. R. Lynd. 2012. Formation and characterization of non-

growth states in *Clostridium thermocellum*: spores and L-forms. *BMC Microbiol*. 12:180. http://dx.doi.org/10.1186/1471-2180-12-180.

Miller, R. A., S. M. Beno, D. J. Kent, L. M. Carroll, N. H. Martin, K. J. Boor, and J. Kovac. 2016. *Bacillus wiedmannii* sp. nov. is a new psychrotolerant and cytotoxic *Bacillus cereus* group species isolated from dairy foods and environments in the USA. *Int. J. Sys. Evol. Microbiol*. 66 (11):4733-4753.

Moreno Switt, A. I., A. D. Andrus, M. L. Ranieri, R. H. Orsi, R. I. H. den Bakker, N. H. Martin, M. Wiedmann, and K. J. Boor. 2014. Genomic comparison of sporeforming bacilli isolated from milk. *BMC Genomics*. 15:26. http://dx.doi.org/10.1186/1471-2164/15/26.

Nakamura, L. K. 1998. *Bacillus pseudomycoides* sp. nov. *Int. J. Sys. Bacteriol*. 48:1031-1035.

Nakamura, L. K. and M. A. Jackson. 1995. Clarification of the taxonomy of *Bacillus mycoides*. *Int. J. Sys. Bacteriol*. 45(1):46-49.

O'Leary, N. A., M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrun, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, and K. D. Pruitt. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*. 44: D733-D745. http://dx.doi.org/10.1093/narl/gkv1189.

Páčová, Z., P. Švec, L. P. Stenfors, M. Vyletělova, and I. Sedláček. 2003. Isolation of the psychrotolerant species *Bacillus weihenstephanensis* from raw cow's milk. *Czech J. Anim. Sci.* 48(2):93-96.

Palma, F., F. Caravajal, M. Jamilena, and D. Garrido. 2014. Contribution of polyamines and other related metabolites to the maintenance of zucchini fruit quality during cold storage. *Plant Physiol. Biochem*. 82:161-171. http://dx.doi.org/10.1016/j.plaphy.2014.06.001.

Redder, P., S. Hausmann, V. Khemici, H. Yasrebi, and P. Linder. 2015. Bacterial versatility requires DEAD-box RNA helicases. *FEMS Microbiol. Reviews*. 39:392-412. http://dx.doi.org/10.1093/femsre/fuv011.

Saleh-Lakha, S., C. G. Leon-Velarde, S. Chen, S. Lee, K. Shannon, M. Fabri, G. Downing, and B. Keown. 2017. A study to assess the numbers and prevalence of *Bacillus cereus* and its toxins in pasteurized fluid milk. *J. Food Prot*. 80(7):1085-1089. http://dx.doi.org/10.4315/0362-028X.JFP-16-521.

Scharl, T. and F. Leisch. 2006. The stochastic QT-cluster algorithm: Evaluation of stability and variance on time-course microarray data. In Alfredo Rizzi and Maurizio Vichi, editiors, Compstat 2006- Proceedings in Computational Statistics. Pages 1015-1022. Physica Verlag, Heidelberg, Germany.

Schindler, T., P. L. Graumann, D. Perl, S. Ma, F. X. Schmid, and M. A. Marahiel. 1999. The family of cold shcok proteins of *Bacillus subtilis*. *J. Biol. Chem*. 274(6):3407-3413.

Soufiane, B. and J.-C. Côte. 2013. *Bacillus weihenstephanensis* characteristics are present in *Bacillus cereus* and *Bacillus mycoides* strains. *FEMS Microbiol. Lett*. 341:127-137. http://dx.doi.org/10.1111/1574-6968.12106.

Spencer, R. C. 2003. *Bacillus anthracis*. *J. Clin. Pathol*. 56(3):182-187.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30(9):1312-1313. http://dx.doi.org/10.1093/bioinformatics/btu033.

Stenfors, L. P. and P. E. Granum. 2001. Psychrotolerant species from the *Bacillus cereus* group are not necessarily *Bacillus weihenstephanensis*. *FEMS Microbiol. Lett*. 197(2):223-228.

Tatusova, T., M. DiCuccio, A. Badretdin, V. Chetvernin, E. P. Nawrocki, L. Zaslavsky, A.

Lomsadze, K. D. Pruitt, M. Borodovsky, and J. Ostell. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*. doi: 10.1093/narlgkw569.

Warda, A. K., R. J. Siezen, J. Boekhorst, M. H. J. Wells-Bennik, A. de Jong, O. P. Kuipers, M. N. Nierop Groot, T. Abee. 2016. Linking *Bacillus cereus* genotypes and carbohydrate utilization capacity. *PLoS ONE* 11(6):e0156796. http://dx.doi.org/10.1371/journal.pone.0156796.

Zhang, Y., J. Chen, C. Feng, L. Zhan, J. Zhang, Y. Li, Y. Yang, H. Chen, Z. Zhang, Y. Zhang, L. Mei, and H. Li. 2017. Quantitative prevalence, phenotypic and genotypic characteristics of *Bacillus cereus* isolated from retail infant foods in China. *Foodborne Pathog Dis*. http://dx.doi.org/10.1089/fpd.2017.2287.

.

CHAPTER 5

CONCLUSIONS

Bacillales associated with spoilage or illnesses are commonly isolated from dairy products, as well as from dairy production and processing environments. The overall goals of this research were to (i) assess the prevalence of *Listeria* and *L. monocytogenes* in small dairy processing facilities, (ii) assist small dairy processors to meet the new requirements of the Food Safety Modernization Act (FSMA) and (iii) understand the relationship between psychrotolerant spoilage organisms (i.e., *Paenibacillus* and the *B. cereus* group) and their genetics. Ultimately, this research provides extensive data that will be useful to dairy stakeholders in developing new strategies to ensure safer and higher quality products.

In the first study, we tested 4, 430 environmental samples for *Listeria*. These samples were taken from 9 different processing facilities monthly for 2 years. Through pulsed field gel electrophoresis (PFGE) and *sigB* allelic typing, we found evidence for persistence in some processing facilities. Furthermore, the data shows that prevalence of *Listeria* and *L. monocytogenes* greatly varies by processing facility. We hypothesize that these differences may be attributed to facility age, quality of sanitary equipment design, and differences in sample collection, among other things. While processors were each trained to collect samples, we were not present for each sample collection and cannot guarantee that instructions were carefully followed for each sampling at each facility. Importantly, artisanal cheese processors were participating voluntarily, with an incentive to sell cheese to Wegmans, a large grocer. We hypothesize that some of these facilities (e.g., Facility J) may have set a goal of not finding Listeria under pressure of the deal with Wegmans. Although Wegmans did not receive the results for each individual facility, the results of this study are thought to be representative of small cheese-processing facilities.

Ultimately, our design and implementation of pathogen environmental monitoring

programs for small processing facilities provides guidance for artisanal processors and other small facilities, as the Food Safety Modernization Act (FSMA) requires that all facilities where ready-to-eat products are exposed to the processing environment have an environmental monitoring program. Importantly, this research provides guidance for validating these pathogen environmental monitoring programs to ensure the programs are as effective as possible, ultimately leading to a safer and more responsible food supply.

We ran the same analyses for both the second and the third study. These analyses ultimately set up future research for screening tools to be used throughout the dairy industry and together, show that at 6°C growth is often stochastic. While *Paenibacillus* and *B. cereus* group isolates are often found in similar environments, the results from the analyses show that these organisms may require different risk assessment and screening strategies. *Paenibacillus* spp. showed varying abilities to grow at 6°C in SMB. Growth patterns were not associated with phylogenetic clades. Interestingly, we did not find any SNPs, genes, or GO terms associated with the ability of an isolate to grow at 6°C in SMB. This is particularly challenging given that we would like to develop a screening tool to more quickly identify *Paenibacillus* that will cause problems for spoilage later in shelf life. Through the ANIb analysis, we also concluded that more analysis is needed to identify many *Paenibacillus* isolates to the species level. We could not name 18 species, and when compared with the 16S rDNA tree, many of these isolates may represent new *Paenibacillus* species. Future research will categorize these isolates to the species level and potentially result in descriptions of new dairy-associated *Paenibacillus* species.

The *B. cereus* group presents unique challenges. Using many of the same genetic analyses, we found many genes and GO terms associated with psychrotolerant *B. cereus* group isolates. Furthermore, phylogenetic clade VI, representing *B. weihenstephanensis* and *B. mycoides* is associated with psychrotolerance. While this is promising for the development of *B. cereus* group screening tools, many isolates were unable to grow in either SMB or BHI broth at 6°C. However, many of these same isolates grew during screenings on BHI agar at the same temperature. Most of

the isolates that grew in BHI were unable to grow in SMB, suggesting that food matrix may play a big role in these screening methods. Future research is needed to more fully understand the interactions of media, temperature, and other stress conditions on the ability of *B. cereus* group isolates to grow.

Overall, the data collected throughout these projects are beneficial to the dairy industry. The first study provided small-processors with experience in environmental sampling plans and the corrective actions required. This will ultimately lead to compliance with FSMA, cleaner processing environments, and a safer food supply. Annecdotal evidence of reduced *L. monocytogenes* prevalence in the processing environment encourages small processors to invest in their environmental sampling programs. In addition, our research on the genetics of psychrotolerant spore-formers ultimately provides the baseline data for the development of screening tools that can be used throughout the dairy value chain to determine most influential contamination areas and to effectively eliminate psychrotolerant spore-formers. Ultimately, future research from these analyses may lead to less food spoilage and food waste.

APPENDIX A

Supplemental Table 2.1: Routine *Listeria* sampling locations by facility

Facility	Routine Sampling Sites [a]
A	Floor: office (3), storage rooms (2), production rooms (3), aging rooms (2) doorway to cooler (1); Drains: bulk tank room (1), production room (2), aging room (1), walk in cooler (1), Equipment: cart wheels (2), underside bulk tank (1), underside vats (2), draining table (1); Other: pipes in aging room (1), aging room ceiling (1), racks in cooler (1), squeegees (2), underside trash cans (1)
C	Floor: entry (1), processing room (2), aging room (3), storage room (1), doorway to aging room (1); Equipment: underside of vat (1), underside of table (2), shelving in aging room (3), brush for floors (1), squeegee (1); Ceilings: processing room (1), aging room (1); Other: drain in processing room (1), sink in processing room (2), underside of trash can (1), wall port (1)
D	Floor: processing room (2), raw milk room (1), farm store (1); Equipment: underside of vat (2), refrigerator (3), cart (1), table (1), draining pans (1), broom (1), brush (1); Other: ceiling of processing room (1), sinks (3), underside of trash cans (1), drain in processing room (1)
E	Floor: entry (2), processing room (2), aging room (6); Drains: processing room (2); Equipment: vats (3), brine tanks (2), tables (2), aging shelves (8), brush/squeegee (1); Other: ceiling of aging rooms (3), walls of aging rooms (3), sink (2), trash can (1), hose (1), ladder (1), paddle bridge (1)
F	Floor: processing room (2), aging room (1), storage (1); Drains: processing room (1), aging room (1), cooler room (2); Equipment: vat (1), cart (1), cheese press (1), brush (2), squeegee (1), aging shelf (2); Other: door frame (2), air ports (1), trash can (1), sink (1), windowsill (1), processing room ceiling (1), foot stool (1)
G	Floor: packaging room (2), can storage room (1), processing room (1), aging room (7), elevator (1), vestibule to barn (1); Drains: milk house (1), can storage room (3), processing room (3), wash room (1), brine room (2), aging rooms (7); Other: hose (1), chiller condensation (1), squeegees (5), wheels to carts and wagons (3), shelves (3), wall of aging room (1)
H	Floor: processing room (3), aging room (5); Drains: processing room (2), milk receiving room (1), aging room (1); Other: catch trough (1), hose (2), aging room wall (3)
I	Floor: processing room (4), office (3), hallway (4), aging room (4); Drains: processing room (3), aging rooms (4); Equipment: exterior of vat (3), exterior of balance tank (1), exterior of brine tub (2), cart, underside of table (6), aging room shelves (4); Other: storage shelves (1), door (2), trash can (2), sink (3), keyboard (1)
J	Floor: threshold (1), processing room (5), packing room (1), cooler (1), aging room (6); Equipment: underside of table (1), cart (1), aging shelf (1); Ceiling: aging room (2); Other: whey sink (1), hose (1), underside of door (1)

[a] The number in parentheses indicates the number of samples taken with that description each month.

APPENDIX B

Appendix B-1: Growth data (CFU/ml) for 28 *Paenibacillus* isolates grown in BHI and SMB at 6 and 10°C.

Isolate ID (Replicate)	Day 0				Day 14				Day 21			
	BHI 6°C	SMB 6°C	BHI 10°C	SMB 10°C	BHI 6°C	SMB 6°C	BHI 10°C	SMB 10°C	BHI 6°C	SMB 6°C	BHI 10°C	SMB 10°C
F4-0085 (1)	7.60E+2	6.00E+2	6.54E+2	6.54E+2	3.68E+7	0.00E+0[a]	1.54E+7	4.60E+5	1.83E+8	0.00E+0[a]	1.24E+7	2.31E+7
F4-0085 (2)	1.58E+3	1.64E+3	1.90E+3	1.29E+3	5.40E+7	1.10E+6	3.15E+7	1.71E+7	7.89E+7	0.00E+0[a]	1.55E+7	3.89E+7
F4-0085 (3)	2.96E+3	3.08E+3	2.88E+3	2.80E+3	4.02E+7	1.44E+6	2.08E+7	3.87E+7	7.28E+7	4.00E+4	1.09E+7	2.24E+7
F4-0126 (1)	9.99E+2	1.20E+3	9.61E+2	6.34E+2	3.28E+7	3.52E+6	1.90E+7	3.91E+7	1.72E+7	2.90E+6	3.00E+6	1.91E+7
F4-0126 (2)	1.10E+3	9.59E+2	1.19E+3	1.06E+3	3.71E+7	4.14E+6	2.39E+7	2.49E+7	2.03E+7	3.20E+6	3.92E+6	1.92E+7
F4-0126 (3)	2.20E+2	6.00E+2	1.84E+2	9.20E+2	2.37E+7	9.39E+5	3.20E+5	3.80E+5	2.14E+7	0.00E+0[a]	3.16E+6	1.29E+7
F4-0134 (1)	4.92E+3	3.52E+3	3.92E+3	3.39E+3	5.41E+7	0.00E+0[a]	3.00E+5	2.02E+6	1.47E+7	0.00E+0[a]	2.58E+6	2.61E+7
F4-0134 (2)	4.18E+3	3.76E+3	4.03E+3	3.50E+3	2.65E+7	2.00E+4	2.00E+4	2.40E+5	1.81E+7	0.00E+0[a]	4.58E+6	2.74E+7
F4-0134 (3)	1.20E+3	1.06E+3	9.61E+2	6.13E+2	1.34E+7	0.00E+0[a]	8.00E+4	4.06E+6	2.50E+7	0.00E+0[a]	3.50E+6	3.28E+7
F4-0152 (1)	3.20E+2	4.80E+2	4.50E+2	4.29E+2	6.25E+6	1.20E+6	1.48E+7	2.11E+7	2.03E+7	0.00E+0[a]	2.94E+7	1.44E+7
F4-0152 (2)	1.80E+2	8.00E+1	2.25E+2	5.31E+2	8.23E+6	0.00E+0[a]	2.00E+7	2.35E+7	3.41E+7	0.00E+0[a]	1.00E+7	2.58E+7
F4-0152 (3)	2.60E+2	6.00E+1	1.23E+2	1.43E+2	1.72E+7	0.00E+0[a]	2.19E+7	1.36E+7	4.61E+7	0.00E+0[a]	3.74E+6	5.95E+7
F4-0242 (1)	3.00E+2	2.00E+1	0.00E+0[a]	4.29E+2	1.32E+6	0.00E+0[a]	4.24E+6	8.36E+7	3.02E+6	0.00E+0[a]	2.19E+6	5.87E+7
F4-0242 (2)	2.00E+1	2.00E+2	2.04E+1	4.09E+1	3.13E+8	0.00E+0[a]	4.74E+7	4.62E+7	3.41E+6	0.00E+0[a]	3.21E+6	3.11E+7
F4-0242 (3)	2.00E+2	6.00E+2	1.64E+2	5.52E+2	7.69E+7	0.00E+0[a]	3.54E+6	5.59E+8	6.51E+6	1.02E+5	3.64E+6	2.40E+7
H3-0280 (1)	0.00E+0[a]	0.00E+0[a]	0.00E+0[a]	0.00E+0[a]	1.20E+5	0.00E+0[a]	1.74E+6	6.87E+6	0.00E+0[a]	0.00E+0[a]	0.00E+0[a]	0.00E+0[a]
H3-0280 (2)	4.00E+1	0.00E+0[a]	6.00E+1	1.40E+2	2.39E+5	2.06E+5	2.76E+5	3.58E+5	1.02E+7	3.12E+6	2.50E+6	1.09E+7
H3-0280 (3)	4.60E+2	2.00E+1	6.00E+2	6.00E+1	3.47E+5	2.86E+5	2.03E+5	4.18E+5	1.08E+7	6.92E+6	2.29E+6	1.12E+7
H3-0287 (1)	2.00E+2	2.00E+1	4.00E+1	1.60E+2	1.73E+5	5.80E+2	2.05E+5	2.32E+5	1.66E+7	0.00E+0[a]	5.67E+6	6.70E+6
H3-0287 (2)	4.00E+1	8.00E+1	0.00E+0[a]	2.00E+2	2.76E+3	2.36E+3	4.07E+5	2.64E+5	4.55E+6	1.79E+5	6.00E+6	6.89E+6
H3-0287 (3)	6.00E+1	1.40E+2	2.00E+1	4.00E+1	8.00E+4	1.62E+3	1.88E+5	8.00E+4	1.31E+7	0.00E+0[a]	5.15E+6	1.17E+7
H3-0464 (1)	4.00E+1	4.00E+1	4.00E+1	0.00E+0[a]	2.22E+3	2.00E+1	1.88E+5	6.00E+2	1.27E+6	0.00E+0[a]	2.74E+6	0.00E+0[a]
H3-0464 (2)	3.80E+2	6.00E+1	2.04E+3	9.99E+1	6.40E+4	2.20E+2	3.99E+5	2.00E+2	3.30E+6	2.10E+4	5.12E+6	6.30E+4
H3-0464 (3)	1.02E+3	2.00E+1	2.00E+2	2.00E+1	8.04E+4	1.60E+2	3.46E+5	4.60E+2	3.48E+6	0.00E+0[a]	4.47E+6	0.00E+0[a]
H3-0465 (1)	2.00E+1	4.00E+1	4.00E+1	8.00E+1	5.67E+4	2.64E+3	1.97E+5	1.98E+5	5.45E+6	0.00E+0[a]	1.57E+7	2.24E+7
H3-0465 (2)	4.00E+1	6.00E+1	4.00E+1	4.00E+1	0.00E+0[a]	0.00E+0[a]	1.41E+8	4.00E+7	1.05E+4	0.00E+0[a]	1.24E+7	2.37E+7

H3-0465 (3)	2.00E+1	2.40E+2	8.00E+1	0.00E+0[a]	5.67E+4	0.00E+0[a]	7.81E+7	2.69E+8	6.45E+6	0.00E+0[a]	1.23E+7	8.39E+6
H7-0433 (1)	1.60E+2	2.00E+1	1.40E+2	2.00E+1	2.39E+5	0.00E+0[a]	1.89E+5	1.30E+5	1.71E+7	0.00E+0[a]	6.27E+6	1.36E+7
H7-0433 (2)	2.40E+2	5.00E+2	1.20E+2	6.00E+1	9.68E+4	0.00E+0[a]	2.29E+5	1.42E+5	1.33E+7	0.00E+0[a]	9.30E+6	8.89E+6
H7-0433 (3)	1.20E+2	7.87E+3	2.00E+2	2.00E+2	3.40E+4	3.40E+2	1.58E+5	4.01E+5	3.59E+6	0.00E+0[a]	7.41E+6	9.31E+6
H7-0443 (1)	0.00E+0[a]	4.00E+1	0.00E+0	8.00E+1	2.55E+6	0.00E+0[a]	3.07E+7	1.68E+7	4.70E+7	0.00E+0[a]	4.11E+6	1.52E+7
H7-0443 (2)	0.00E+0[a]	0.00E+0[a]	4.60E+2	6.00E+1	1.55E+6	0.00E+0[a]	1.59E+7	1.47E+7	5.32E+7	0.00E+0[a]	3.31E+6	1.36E+7
H7-0443 (3)	0.00E+0[a]	0.00E+0[a]	2.00E+1	0.00E+0[a]	5.52E+5	0.00E+0[a]	1.72E+7	1.48E+7	9.23E+7	4.09E+4	2.51E+6	2.25E+7
H7-0604 (1)	1.84E+2	1.23E+2	1.43E+2	3.88E+2	2.40E+5	0.00E+0[a]	2.51E+7	1.80E+7	1.30E+7	0.00E+0[a]	9.79E+6	1.49E+7
H7-0604 (2)	1.64E+2	1.02E+2	2.04E+2	6.13E+1	8.00E+4	0.00E+0[a]	2.27E+7	3.74E+6	1.31E+7	0.00E+0[a]	5.47E+6	4.00E+6
H7-0604 (3)	6.74E+2	0.00E+0[a]	6.13E+1	4.09E+1	2.40E+5	0.00E+0[a]	2.07E+7	6.87E+6	2.05E+7	0.00E+0[a]	5.38E+6	7.08E+6
H7-0694 (1)	8.18E+1	1.43E+2	1.64E+2	8.18E+1	7.60E+6	9.99E+4	1.86E+7	6.23E+6	1.12E+7	6.00E+4	9.93E+6	9.15E+6
H7-0694 (2)	2.04E+1	6.13E+1	2.04E+1	0.00E+0[a]	3.72E+6	0.00E+0[a]	2.73E+7	6.66E+6	4.59E+7	2.36E+6	4.18E+6	1.12E+7
H7-0694 (3)	0.00E+0[a]	0.00E+0[a]	2.04E+1	0.00E+0[a]	7.31E+6	0.00E+0[a]	3.79E+7	3.80E+6	1.68E+7	0.00E+0[a]	6.80E+6	3.74E+6
H7-0710 (1)	3.27E+2	1.64E+2	1.02E+2	1.64E+2	2.54E+7	0.00E+0[a]	4.78E+7	3.90E+6	9.96E+7	2.42E+6	2.33E+7	8.28E+6
H7-0710 (2)	2.56E+4	1.71E+4	2.24E+4	7.56E+2	3.26E+6	2.88E+6	3.84E+7	7.08E+6	3.80E+6	6.02E+6	1.65E+7	7.98E+6
H7-0710 (3)	2.04E+1	6.13E+1	0.00E+0[a]	4.09E+1	2.02E+7	4.00E+4	4.83E+7	3.10E+6	9.11E+7	4.00E+4	1.94E+7	8.89E+6
H7-0713 (1)	4.09E+1	2.04E+1	0.00E+0[a]	0.00E+0[a]	1.66E+6	0.00E+0[a]	1.97E+7	7.90E+6	0.00E+0[a]	8.71E+6	1.09E+8	6.16E+6
H7-0713 (2)	0.00E+0[a]	6.13E+1	0.00E+0[a]	0.00E+0[a]	2.39E+6	0.00E+0[a]	3.58E+7	5.01E+6	1.04E+8	0.00E+0[a]	1.25E+7	9.64E+6
H7-0713 (3)	0.00E+0[a]	0.00E+0[a]	0.00E+0[a]	6.00E+1	2.31E+6	0.00E+0[a]	4.70E+7	4.01E+6	9.59E+7	4.70E+4	1.21E+7	4.26E+6
H7-0718 (1)	6.00E+1	8.00E+1	8.00E+1	9.99E+1	1.60E+7	0.00E+0[a]	7.72E+6	4.09E+5	4.77E+7	2.00E+4	1.05E+7	4.43E+7
H7-0718 (2)	0.00E+0[a]	2.60E+2	2.00E+2	1.80E+2	1.46E+7	0.00E+0[a]	1.11E+7	1.10E+6	3.86E+6	0.00E+0[a]	8.16E+6	4.30E+7
H7-0718 (3)	2.60E+2	3.80E+2	2.60E+2	2.20E+2	1.98E+6	4.09E+4	1.23E+7	9.61E+5	2.77E+7	1.86E+6	5.62E+6	6.48E+7
H7-0744 (1)	0.00E+0[a]	0.00E+0[a]	0.00E+0[a]	0.00E+0[a]	2.25E+2	4.09E+2	1.34E+4	4.29E+2	3.77E+4	2.66E+3	3.72E+4	3.00E+3
H7-0744 (2)	0.00E+0[a]	0.00E+0[a]	0.00E+0[a]	0.00E+0[a]	8.18E+1	2.66E+2	7.36E+2	1.49E+3	9.13E+3	2.25E+3	1.47E+5	3.34E+5
H7-0744 (3)	0.00E+0[a]	0.00E+0[a]	0.00E+0[a]	0.00E+0[a]	1.23E+2	0.00E+0[a]	2.94E+3	5.12E+2	1.64E+2	5.93E+2	1.37E+5	2.33E+5
H7-0918 (1)	8.00E+1	6.00E+1	9.99E+1	1.20E+2	5.92E+6	0.00E+0[a]	3.21E+6	1.34E+7	3.14E+7	0.00E+0[a]	1.54E+6	4.11E+7
H7-0918 (2)	8.40E+1	7.35E+1	1.68E+2	8.40E+1	3.27E+6	0.00E+0[a]	4.87E+6	1.63E+7	3.21E+7	0.00E+0[a]	2.76E+6	3.61E+7
H7-0918 (3)	1.20E+2	2.20E+2	6.00E+1	6.00E+1	4.87E+6	0.00E+0[a]	3.90E+6	3.08E+7	3.56E+7	0.00E+0[a]	4.90E+7	4.32E+7
H8-0069 (1)	2.00E+1	6.00E+1	4.00E+1	8.00E+1	1.29E+6	0.00E+0[a]	2.39E+7	2.28E+7	1.94E+7	0.00E+0[a]	2.80E+7	3.50E+7
H8-0069 (2)	0.00E+0[a]	6.00E+1	1.40E+2	6.00E+1	1.86E+6	0.00E+0[a]	3.24E+7	2.46E+7	1.73E+7	0.00E+0[a]	1.93E+7	3.87E+7
H8-0069 (3)	2.00E+1	4.00E+1	6.00E+1	1.20E+2	1.90E+6	0.00E+0[a]	2.46E+7	1.26E+7	1.69E+7	7.80E+5	1.43E+7	1.00E+7

H8-0147 (1)	1.80E+2	1.80E+2	9.99E+1	2.80E+2	1.27E+6	0.00E+0[a]	2.74E+7	1.25E+7	2.13E+7	2.00E+4	1.51E+7	2.60E+7
H8-0147 (2)	4.00E+1	0.00E+0[a]	2.20E+2	8.00E+1	0.00E+0[a]	0.00E+0[a]	2.11E+7	1.23E+7	1.29E+7	0.00E+0[a]	1.97E+7	1.10E+7
H8-0147 (3)	1.40E+2	7.55E+3	1.80E+2	2.20E+2	2.04E+4	0.00E+0[a]	1.65E+7	1.07E+7	1.79E+7	0.00E+0[a]	2.50E+7	2.01E+7
H8-0175 (1)	2.08E+3	1.70E+3	1.58E+3	2.02E+3	8.97E+6	0.00E+0[a]	3.93E+7	2.60E+7	6.29E+7	1.80E+5	7.98E+6	2.09E+7
H8-0175 (2)	1.85E+4	1.63E+4	1.90E+4	1.79E+4	5.06E+7	3.45E+6	2.17E+7	1.39E+7	6.14E+7	1.52E+7	9.67E+6	2.05E+7
H8-0175 (3)	2.66E+3	2.36E+3	2.66E+3	2.94E+3	1.09E+7	3.88E+5	2.48E+7	2.11E+7	8.00E+7	2.76E+6	1.09E+7	2.10E+7
H8-0237 (1)	9.19E+2	0.00E+0[a]	0.00E+0[a]	1.04E+3	3.00E+6	0.00E+0[a]	4.45E+7	4.74E+6	1.05E+7	0.00E+0[a]	1.47E+7	2.74E+7
H8-0237 (2)	5.02E+3	1.83E+4	6.02E+3	3.84E+3	7.26E+7	4.11E+6	3.26E+7	2.05E+7	6.06E+7	8.87E+6	2.18E+7	2.04E+7
H8-0237 (3)	6.63E+3	4.68E+3	5.88E+3	7.37E+3	4.21E+7	1.23E+5	2.09E+7	1.87E+7	3.88E+7	5.60E+5	1.16E+7	1.68E+7
J3-0153 (1)	1.26E+3	1.02E+3	8.00E+2	1.20E+3	2.07E+7	0.00E+0[a]	2.18E+7	0.00E+0[a]	1.06E+8	1.60E+5	1.22E+7	9.56E+6
J3-0153 (2)	1.36E+3	1.20E+3	1.18E+3	9.39E+2	3.75E+7	4.09E+4	1.75E+7	2.04E+4	1.07E+7	2.00E+4	1.04E+7	6.60E+5
J3-0153 (3)	1.16E+3	1.10E+3	9.19E+2	8.59E+2	1.92E+7	0.00E+0[a]	2.43E+7	4.09E+4	9.19E+5	2.00E+5	4.57E+7	8.00E+5
J3-0155 (1)	3.53E+5	3.40E+5	3.22E+5	1.99E+5	3.67E+7	1.19E+7	2.44E+7	4.68E+7	5.73E+7	6.04E+7	2.30E+7	3.47E+7
J3-0155 (2)	2.25E+5	1.34E+5	4.28E+3	4.06E+3	1.45E+7	2.04E+6	3.88E+7	1.07E+7	4.88E+7	1.93E+8	7.72E+7	2.60E+7
J3-0155 (3)	3.60E+2	1.20E+2	2.40E+2	1.80E+2	1.66E+6	0.00E+0[a]	3.41E+7	2.12E+7	5.76E+7	1.05E+7	1.61E+7	2.10E+7
J3-0159 (1)	2.96E+3	2.84E+3	3.08E+3	2.76E+3	5.08E+7	9.29E+6	1.07E+7	9.48E+6	1.83E+7	1.25E+7	3.40E+6	1.14E+7
J3-0159 (2)	2.38E+3	2.34E+3	2.06E+3	1.38E+3	4.81E+7	1.19E+6	7.39E+6	1.03E+7	2.90E+7	7.28E+6	3.96E+6	1.60E+7
J3-0159 (3)	8.00E+1	4.00E+2	2.00E+2	5.60E+2	3.05E+7	2.45E+5	1.28E+7	9.78E+6	3.10E+7	8.16E+6	3.74E+6	6.70E+6
R5-0636 (1)	3.20E+2	4.00E+2	4.60E+2	3.80E+2	3.02E+7	6.16E+6	1.46E+8	3.06E+6	1.27E+8	9.81E+5	3.15E+8	1.17E+06
R5-0636 (2)	6.00E+2	5.40E+2	5.20E+2	5.20E+2	1.04E+6	5.80E+5	2.09E+8	1.75E+8	3.33E+6	1.31E+8	1.28E+8	1.30E+8
R5-0636 (3)	3.80E+2	3.60E+2	3.80E+2	4.20E+2	9.99E+4	4.60E+5	1.72E+8	4.04E+6	1.23E+5	4.29E+5	3.37E+8	4.81E+6
R5-0883 (1)	2.40E+2	2.80E+2	2.20E+2	8.00E+1	1.76E+7	0.00E+0[a]	9.35E+7	1.32E+7	2.08E+8	0.00E+0[a]	1.23E+8	1.34E+7
R5-0883 (2)	4.00E+1	8.00E+1	4.00E+1	2.00E+1	0.00E+0[a]	8.59E+5	1.69E+8	1.40E+6	8.18E+4	6.51E+7	1.42E+8	1.62E+7
R5-0883 (3)	3.40E+2	3.20E+2	3.40E+2	1.60E+2	2.28E+7	9.99E+4	2.19E+8	1.51E+7	1.15E+8	7.40E+7	3.72E+8	3.79E+7
R5-0923 (1)	4.00E+1	6.00E+1	6.00E+1	2.00E+1	1.88E+7	0.00E+0[a]	1.35E+8	1.32E+6	2.88E+8	0.00E+0[a]	8.99E+5	1.34E+7
R5-0923 (2)	1.20E+2	1.60E+2	1.40E+2	2.00E+1	1.69E+7	1.20E+5	1.32E+8	3.58E+8	1.57E+8	1.12E+8	3.23E+8	1.34E+8
R5-0923 (3)	9.99E+1	3.00E+2	2.00E+2	1.40E+2	2.80E+5	0.00E+0[a]	1.28E+8	1.57E+8	3.52E+7	0.00E+0[a]	1.47E+8	1.32E+8
R5-0937 (1)	2.00E+2	6.00E+1	6.00E+1	2.40E+2	7.00E+5	0.00E+0[a]	6.75E+6	6.34E+6	4.79E+7	1.71E+6	4.31E+6	1.10E+7
R5-0937 (2)	2.00E+1	2.00E+1	0.00E+0[a]	6.00E+1	0.00E+0[a]	0.00E+0[a]	3.47E+7	4.79E+6	2.98E+6	7.90E+6	8.31E+7	4.25E+6
R5-0937 (3)	4.00E+1	0.00E+0[a]	2.00E+1	0.00E+0[a]	4.00E+4	0.00E+0[a]	1.42E+8	2.34E+6	4.80E+6	8.18E+4	1.24E+8	5.78E+6

[a]These values (0.00E+0) are below the limit of detection. The detection limit for this study was 10 CFU/ml. As such, the true values of these counts are < 10 CFU/ml.

Appendix B-2: Growth data from in depth analysis for *P. odorifer* isolates grown in SMB at 6°C where day 0 is given in CFU/ml and days 14 and 21 are expressed as log CFU/ml change from day 0.

	In depth replicate1			In depth replicate 2			In depth replicate 3		
P. odorifer strain	Day 0	Day 14	Day 21	Day 0	Day 14	Day 21	Day 0	Day 14	Day 21
FSL F4-0085	6.00E+02	≤ -1.78*	≤ -1.78*	1.64E+03	2.83	≤ -2.49*	3.08E+03	2.67	1.11
FSL F4-0126	1.20E+03	3.47	3.38	9.59E+02	3.64	3.52	6.00E+02	3.19	≤ -1.78*
FSL F4-0134	3.52E+03	≤ -2.55*	≤ -2.55*	3.76E+03	0.73	≤ -2.58*	1.06E+03	≤ -2.03*	≤ -2.03*
FSL F4-0152	4.80E+02	3.4	≤ -1.68*	8.00E+01	≤ -0.9*	≤ -0.9*	6.00E+01	≤ -0.78*	≤ -0.78*
FSL F4-0242	2.00E+01	≤ -0.3*	≤ -0.3*	2.00E+02	≤ -1.3*	≤ -1.3*	6.00E+02	≤ -1.78*	2.23
FSL H3-0280	0.00E+00[a]	0[a]	0[a]	0.00E+00[a]	0[a]	0[a]	2.00E+01	4.16	5.54
FSL H3-0287	2.00E+01	1.46	≤ -0.3*	8.00E+01	1.47	3.35	1.40E+02	1.06	≤ -1.15*
FSL H3-0464	4.00E+01	-0.3*	≤ -0.6*	6.00E+01	0.56	2.54	2.00E+01	0.9	≤ -0.3*
FSL H3-0465	4.00E+01	1.82	≤ -0.6*	6.00E+01	≤ -0.78*	≤ -0.78*	2.40E+02	≤ -1.38*	≤ -1.38*
FSL H7-0433	2.00E+01	≤ -0.3*	≤ -0.3*	5.00E+02	≤ -1.7*	≤ -1.7*	7.87E+03	-1.36	≤ -2.9*
FSL H7-0604	1.23E+02	≤ -1.09*	≤ -1.09*	1.02E+02	≤ -1.01*	≤ -1.01*	0.00E+00[a]	0[a]	0[a]
FSL H7-0694	1.43E+02	2.84	2.62	6.13E+01	≤ -0.79*	4.59	0.00E+00[a]	0[a]	0[a]
FSL H7-0713	2.04E+01	≤ -0.31*	5.63	6.13E+01	≤ -0.79*	≤ -0.79*	0.00E+00[a]	0[a]	4.67
FSL H7-0718	8.00E+01	≤ -0.9*	2.4	2.60E+02	≤ -1.41*	≤ -1.41*	3.80E+02	2.03	3.69
FSL H7-0918	6.00E+01	≤ -0.78*	≤ -0.78*	7.30E+01	≤ -0.87*	≤ -0.87*	2.20E+02	≤ -1.34*	≤ -1.34*
FSL H8-0069	6.00E+01	≤ -0.78*	≤ -0.78*	6.00E+01	≤ -0.78*	≤ -0.78*	4.00E+01	≤ -0.16*	4.29
FSL H8-0147	1.80E+02	≤ -1.26*	2.05	0.00E+00[a]	0[a]	0[a]	7.55E+03	≤ -2.88*	≤ -2.88*
FSL H8-0175	1.70E+03	-3.23	2.02	1.63E+04	2.33	2.97	2.36E+03	2.22	3.07
FSL H8-0237	0.00E+00[a]	0[a]	0[a]	1.83E+04	2.35	2.69	4.68E+03	1.42	2.08
FSL J3-0153	1.02E+03	≤ -2.01*	2.2	1.20E+03	1.53	1.22	1.10E+03	≤ -2.04*	2.26
FSL J3-0155	3.40E+05	1.54	2.25	1.34E+05	1.18	3.16	1.20E+02	≤ -1.08*	4.94
FSL J3-0159	2.84E+03	3.51	3.64	2.34E+03	2.71	3.49	4.00E+02	2.79	4.31
FSL R5-0636	4.00E+02	4.19	3.39	5.40E+02	3.03	5.38	3.60E+02	3.11	3.08
FSL R5-0883	2.80E+02	≤ -1.45*	≤ -1.45*	8.00E+01	4.03	5.91	3.20E+02	2.49	5.36
FSL R5-0937	6.00E+01	≤ -0.78*	4.45	2.00E+01	≤ -0.3*	5.6	0.00E+00[a]	0[a]	0[a]

*These values were below the limit of detection (count of 0, meaning < 10 CFU/ml). Because we cannot assign a true reduction to these counts, they are marked as "at least" (≤) this given reduction.

[a]These values (0.00E+0 and 0) are below the limit of detection. The limit of detection in this study was 10 CFU/ml, as such, these values are < 10 CFU/ml.

Appendix B-3: Growth data from initial screen for *P. odorifer* isolates grown in SMB at 6°C where day 0 is given in CFU/ml and days 14 and 21 are expressed as log CFU/ml change from day 0.

	Screen Replicate 1			Screen Replicate 2			Screen Replicate 3[a]		
P. odorifer strain	Day 0	Day 14	Day 21	Day 0	Day 14	Day 21	Day 0	Day 14	Day 21
FSL F4-0085	5.37E+02	-1.18	-0.16	5.16E+02	0.58	3.52	-	-	-
FSL F4-0126	1.00E+01	0.99	3.76	7.67E+01	3.57	5.61	-	-	-
FSL F4-0134	2.88E+03	1.85	3.41	5.62E+01	2.97	5.57	-	-	-
FSL F4-0152	1.67E+04	-0.59	2.25	8.18E+01	2	4.79	-	-	-
FSL F4-0242	1.16E+03	1.08	3.41	1.39E+04	3.96	6.17	-	-	-
FSL H3-0280	3.17E+02	2.22	4.11	1.24E+03	3.17	4.45	-	-	-
FSL H3-0287	6.83E+03	3.02	3.69	1.02E+01	2.97	6.18	-	-	-
FSL H3-0464	2.05E+02	3.51	4.9	1.79E+02	2.78	5.05	-	-	-
FSL H3-0465	5.65E+04	2.26	3.03	1.33E+02	1.81	3.96	-	-	-
FSL H7-0433	7.55E+03	3.31	3.76	2.25E+02	2.54	4.49	-	-	-
FSL H7-0604	9.86E+02	2.29	4.48	1.89E+02	2.69	4.82	-	-	-
FSL H7-0694	1.31E+03	3.25	4.21	1.55E+03	3.81	4.5	-	-	-
FSL H7-0713	5.37E+02	-0.91	-1.12	6.43E+03	3.34	3.73	-	-	-
FSL H7-0718	1.50E+03	-1.56	-2.17	1.21E+03	3.59	4.45	-	-	-
FSL H7-0918	2.68E+03	1.66	3.31	1.03E+03	3.61	4.41	-	-	-
FSL H8-0069	7.00E+02	3.11	4.32	1.99E+03	3.91	4.33	-	-	-
FSL H8-0147	3.68E+02	-1.57	-1.56	1.78E+03	-1.31	-1.24	-	-	-
FSL H8-0175	1.31E+03	1.24	3.22	2.35E+02	2.58	4.56	-	-	-
FSL H8-0237	1.18E+02	-1.47	-1.47	8.13E+02	4.36	4.64	4.81E+02	4.77	5.04
FSL J3-0153	7.51E+02	3.14	4.26	6.82E+02	4.72	4.73	4.04E+02	4.4	4.91
FSL J3-0155	5.98E+02	4.26	4.73	4.91E+02	4.99	4.84	-	-	-
FSL J3-0159	2.40E+02	5.13	5.19	1.02E+02	5.46	5.55	-	-	-
FSL R5-0636	6.80E+03	1.48	3.6	2.50E+03	2.4	3.96	-	-	-
FSL R5-0883	5.51E+03	0.63	2.8	2.38E+03	1.36	3.75	-	-	-
FSL R5-0937	2.04E+01	-0.3	0.18	5.98E+03	2.1	3.74	5.11E+01	2.6	5.46

[a] Three isolates were tested in triplicate, but all others were tested in duplicate.

Supplemental Table 3.1: List of 101 isolates tested in the initial screen and associated growth data for all isolates in Skim Milk Broth (SMB) at 6°C

Isolate	Genus	species	Growth?*	Day 14 Growth	Day 21 Growth	rpoB AT	rpoB AT frequency	Comments
FSL R5-0172	*Bacillus*	*aerophilus s.l.*	-	-1.85	-1.85	206	2 times	
FSL R5-0226	*Bacillus*	*aerophilus s.l.*	-	-1.56	-1.56	69	4 times	
FSL R5-0860	*Bacillus*	*cereus*	-	-1.57	-1.60	158	137 times	
FSL H7-0588	*Bacillus*	*cereus s.l.*	-	-2.75	-2.75	59	26 times	
FSL H7-0967	*Bacillus*	*cereus s.l.*	-	-1.44	-1.44	92	5 times	
FSL H8-0049	*Bacillus*	*cereus s.l.*	-	-1.36	-1.36	61	4 times	
FSL J3-0113	*Bacillus*	*cereus s.l.*	-	-1.87	-1.87	417	NA	Isolate from Trmčić et al., 2015
FSL R5-0673	*Bacillus*	*cereus s.l.*	-	-1.80	-1.80	246	4 times	
FSL H8-0454	*Bacillus*	*cf. badius*	-	-1.57	-1.75	116	1 time	
FSL H8-0521	*Bacillus*	*cf. cecembensis*	-	-2.02	-2.02	145	6 times	
FSL R5-0597	*Bacillus*	*cf. megaterium*	-	-3.13	-3.14	248	1 time	
FSL H7-0305	*Bacillus*	*clausii*	-	-1.52	-1.52	55	2 times	
FSL R5-0254	*Bacillus*	*firmus*	-	-1.20	-1.20	209	3 times	
FSL R7-0078	*Bacillus*	*horikoshii*	-	-1.59	-1.68	174	2 times	
FSL H7-0968	*Bacillus*	*licheniformis*	-	-1.59	-1.59	1	134 times	
FSL J3-0143	*Bacillus*	*licheniformis*	-	-1.45	-1.45	1	134 times	
FSL R5-0393	*Bacillus*	*licheniformis s.l.*	-	-0.50	-0.50	215	3 times	
FSL R5-0402	*Bacillus*	*licheniformis s.l.*	-	-0.50	-0.50	169	2 times	
FSL F4-0108	*Bacillus*	*licheniformis s.l.*	-	-0.31	-0.31	9	11 times	
FSL H7-0358	*Bacillus*	*licheniformis s.l.*	-	-0.35	-0.36	6	35 times	
FSL R5-0395	*Bacillus*	*mojavensis*	-	-0.16	-0.16	217	2 times	
FSL J3-0137	*Bacillus*	*muralis*	-	-2.01	-2.01	512	NA	Isolate from Trmčić et al., 2015
FSL R7-0271	*Bacillus*	*nealsonii*	-	-1.20	-1.22	191	1 time	
FSL H7-0356	*Bacillus*	*pumilus*	-	-0.99	-0.99	62	6 times	
FSL R5-0176	*Bacillus*	*pumilus*	-	-1.92	-1.92	47	5 times	
FSL R5-0403	*Bacillus*	*pumilus*	-	-2.15	-2.15	68	9 times	
FSL R5-0231	*Bacillus*	*safensis*	-	-2.18	-2.18	140	6 times	
FSL R5-0257	*Bacillus*	*safensis*	-	-0.59	-0.59	210	2 times	

FSL R5-0420	*Bacillus*	*safensis*	-	-2.11	-2.11	222	3 times	
FSL R5-0849	*Bacillus*	sp.	-	-2.47	-2.47	34	3 times	
FSL H7-0434	*Bacillus*	*subtilis s.l.*	-	-0.01	-0.01	66	2 times	
FSL R5-0428	*Bacillus*	*subtilis s.l.*	-	-0.30	0.00	227	2 times	
FSL H7-0432	*Bacillus*	*subtilis s.l.*	-	-1.00	-1.00	65	6 times	
FSL H8-0102	*Bacillus*	*weihenstephanensis*	-	-1.41	-1.41	75	23 times	
FSL J3-0123	*Bacillus*	*weihenstephanensis*	-	-1.65	-1.65	513	NA	Isolate from Trmčić et al., 2015
FSL J3-0124	*Bacillus*	*weihenstephanensis*	Stochastic	1.06	1.90	3	19 times	
FSL R5-0572	*Lysinibacillus*	sp.	-	-1.15	-1.15	242	1 time	
FSL R5-0430	*Oceanobacillus*	sp.	-	-1.89	-1.89	268	1 time	
FSL R5-0808	*Paenibacillus*	*glucanolyticus*	-	-1.92	-1.92	159	6 times	
FSL R5-0817	*Paenibacillus*	*glucanolyticus*	-	0.00	0.55	159	6 times	
FSL R5-0527	*Paenibacillus*	*macerans*	-	0.00	0.12	238	1 time	
FSL H8-0147	*Paenibacillus*	*odorifer*	-	-1.44	-1.40	40	5 times	
FSL F4-0126	*Paenibacillus*	*odorifer*	+	2.28	4.69	13	21 times	
FSL F4-0134	*Paenibacillus*	*odorifer*	+	2.41	4.49	16	6 times	
FSL F4-0152	*Paenibacillus*	*odorifer*	+	0.71	3.52	19	20 times	
FSL F4-0242	*Paenibacillus*	*odorifer*	+	2.52	4.79	25	19 times	
FSL H3-0280	*Paenibacillus*	*odorifer*	+	2.70	4.28	27	79 times	
FSL H3-0287	*Paenibacillus*	*odorifer*	+	3.00	4.94	2	52 times	
FSL H3-0305	*Paenibacillus*	*odorifer*	+	2.48	4.95	38	3 times	
FSL H3-0464	*Paenibacillus*	*odorifer*	+	3.15	4.98	46	5 times	
FSL H3-0465	*Paenibacillus*	*odorifer*	+	2.04	3.50	50	8 times	
FSL H7-0433	*Paenibacillus*	*odorifer*	+	2.93	4.13	36	13 times	
FSL H7-0604	*Paenibacillus*	*odorifer*	+	2.49	4.65	74	5 times	
FSL H7-0694	*Paenibacillus*	*odorifer*	+	3.53	4.36	35	16 times	
FSL H7-0918	*Paenibacillus*	*odorifer*	+	2.64	3.86	88	3 times	
FSL H8-0069	*Paenibacillus*	*odorifer*	+	3.51	4.33	98	4 times	
FSL H8-0175	*Paenibacillus*	*odorifer*	+	1.91	3.89	107	3 times	
FSL J3-0153	*Paenibacillus*	*odorifer*	+	4.09	4.63	46	5 times	
FSL J3-0155	*Paenibacillus*	*odorifer*	+	4.63	4.79	35	16 times	

FSL J3-0159	*Paenibacillus*	*odorifer*	+	5.30	5.37	7	11 times	
FSL R5-0636	*Paenibacillus*	*odorifer*	+	1.94	3.78	180	4 times	
FSL R5-0883	*Paenibacillus*	*odorifer*	+	1.00	3.28	27	79 times	
FSL F4-0077	*Paenibacillus*	*odorifer*	Stochastic	-1.11	-0.63	2	52 times	
FSL F4-0085	*Paenibacillus*	*odorifer*	Stochastic	-1.18	-0.16	4	9 times	
FSL H7-0713	*Paenibacillus*	*odorifer*	Stochastic	1.22	1.31	33	9 times	
FSL H7-0718	*Paenibacillus*	*odorifer*	Stochastic	1.02	1.14	32	16 times	
FSL H8-0237	*Paenibacillus*	*odorifer*	Stochastic	2.56	2.74	15	112 times	
FSL R5-0937	*Paenibacillus*	*odorifer*	Stochastic	1.47	3.13	21	28 times	
FSL R5-0923	*Paenibacillus*	sp.	-	-0.86	-0.34	167	8 times	Unknown species #6- most closely related to *P. odorifer*
FSL A5-0030	*Paenibacillus*	sp.	+	3.79	4.99	179	5 times	Unknown species #14- most closely related to *P. peoriae*
FSL H7-0710	*Paenibacillus*	sp.	+	2.48	3.40	81	6 times	Unknown species #5- closely related to *P. odorifer*
FSL H8-0551	*Paenibacillus*	sp.	+	3.47	4.57	157	8 times	Unknown species #14- most closely related to *P. peoriae*
FSL J3-0120	*Paenibacillus*	sp.	+	3.66	4.97	340	NA	Isolate from Trmčić et al., 2015; Unknown species #14- most closely related to *P. peoriae*
FSL R7-0131	*Paenibacillus*	sp.	+	1.18	3.68	179	5 times	Unknown species #14- most closely related to *P. peoriae*
FSL J3-0122	*Paenibacillus*	sp.	Stochastic	2.34	2.01	23	35 times	Unknown species #17- most closely related to *P. amylolyticus*
FSL R5-0765	*Paenibacillus*	sp.	-	-1.60	-1.59	261	4 times	Unknown species #17- most closely related to *P. xylanexedens*
FSL A5-0031	*Paenibacillus*	sp.	-	-0.31	-0.57	511	NA	Isolate from Trmčić et al., 2015; Unknown species #3- most closely related to *P. endophyticus*
FSL F4-0100	*Paenibacillus*	sp.	-	0.00	0.00	8	3 times	Unknown species #12- most closely related to *P. glucanolyticus* and *P. lautus*
FSL H7-0326	*Paenibacillus*	sp.	-	-0.15	-0.16	57	1 time	Unknown species #15- most closely related to *P. urinalis* and *P. provencensis*

FSL H7-0331	*Paenibacillus*	sp.	-	0.00	-0.62	58	5 times	Unknown species #2- most closely related to *P. vulneris* and *P. rigui*
FSL H7-0692	*Paenibacillus*	sp.	-	-0.09	-0.07	80	4 times	Unknown species #17- most closely related to *P. xylanexedens*
FSL H8-0246	*Paenibacillus*	sp.	-	-0.54	-0.54	108	5 times	Unknown species #16- most closely related to *P. xylanexedens*
FSL H8-0548	*Paenibacillus*	sp.	-	-1.00	-1.00	156	1 time	Unknown species #4- most closely related to *P. castaneae* and *P. endophyticus*
FSL R5-0378	*Paenibacillus*	sp.	-	-2.35	-2.48	214	1 time	Unknown species #13- most closely related to *P. rhizosphoreae*
FSL R5-0490	*Paenibacillus*	sp.	-	-1.46	-1.46	93	3 times	Unknown species #1- new species; no 16S sister clades
FSL R7-0333	*Paenibacillus*	sp.	Stochastic	0.00	-0.79	201	3 times	Unknown species #9- closely related to *P. odorifer*
FSL F4-0087	*Paenibacillus*	sp.	+	1.72	3.05	5	1 time	Unknown species #18- most closely related to *P. taichungensis*
FSL H7-0443	*Paenibacillus*	sp.	+	2.68	4.32	30	12 times	Unknown species #6- closely related to *P. odorifer*
FSL H7-0744	*Paenibacillus*	sp.	+	0.65	2.21	41	3 times	Unknown species #10- closely related to *P. odorifer*
FSL H8-0259	*Paenibacillus*	sp.	+	1.25	2.90	109	1 time	Unknown species #7- new species; most closely related to *P. wynnii*
FSL R7-0269	*Paenibacillus*	sp.	+	1.43	0.70	163	9 times	Unknown species #8- closely related to *P. odorifer*
FSL R7-0277	*Paenibacillus*	sp.	+	1.71	1.83	45	3 times	Unknown species #9- closely related to *P. odorifer*
FSL R7-0321	*Paenibacillus*	sp.	+	0.17	3.15	199	3 times	Unknown species #14- most closely related to *P. peoriae*
FSL R7-0337	*Paenibacillus*	sp.	+	2.99	3.41	202	3 times	Unknown species #8- closely related to *P. odorifer*
FSL F4-0260	*Paenibacillus*	sp.	Stochastic	1.47	2.37	29	3 times	Unknown species #17- most closely related to *P. amylolyticus*
FSL R7-0273	*Paenibacillus*	sp.	Stochastic	-0.61	0.12	193	1 time	Unknown species #11- new

								species; most closely related to *P. wynnii*
FSL H8-0459	*Psychrobacillus*	sp.	-	-2.50	-2.50	119	3 times	
FSL H8-0466	*Psychrobacillus*	sp.	-	-3.13	-3.13	280	1 time	
FSL R5-0425	*Solibacillus*	sp.	-	-0.93	-0.93	226	2 times	
FSL H7-0596	*Viridibacillus*	sp.	+	5.14	5.48	73	18 times	
FSL H8-0123	*Viridibacillus*	sp.	Stochastic	0.95	0.88	73	18 times	

*Growth Symbols:

- represents isolates that were unable to grow at 6°C in SMB for any replicate

+ represents isolates that were able to grow at 6°C in SMB for all replicates

Supplemental Table 3.2: Twelve protein domains previously associated with growth at low temperatures used for Hidden Markov Model (HMM) analyses

Query	Accession	Description	Length
Caps_synth_CapC	PF1402.5	Capsule biosynthesis CapC	119 aa
CSD	PF00313.21	Cold-shock DNA-binding domain	66 aa
DEADboxA	PF12343.7	Cold shock protein DEAD box A	69 aa
DEAD	PF00270.28	DEAD/DEAH box helicase	176 aa
DnaJ	PF00226.30	DnaJ domain	63 aa
FA_desaturase_2	PF03405.13	Fatty acid desaturase	326 aa
FA_desaturase	PF00487.23	Fatty acid desaturase	254 aa
FA_hydroxylase	PF04116.12	Fatty acid hydroxylase superfamily	133 aa
LtrA	PF06772.10	Bacterial low temperature requirement A protein (LtrA)	353 aa
Peptidase_S11	PF00768.19	D-alanyl-D-alanine carboxypeptidase	241 aa
RecA	PF00154.20	recA bacterial DNA recombination protein	263 aa
YdjO	PF14169.5	Cold-inducible protein YdjO	59 aa

Supplemental Table 3.3: Growth data for 28 *Paenibacillus* isolates (25 *P. odorifer* and 3 closely related) in four media/temperature combinations

	Bacterial Counts (log CFU)*							
	10°C BHI		6°C BHI		10°C SMB		6°C SMB	
Isolate	Day 14	Day 21	Day 14	Day 21	Day 14	Day 21	Day 14	Day 21
FSL F4-0085	4.1	3.9	4.5	4.8	3.7	4.3	0.91	-1.6
FSL F4-0126	3.9	3.7	4.7	4.5	3.9	4.3	3.4	1.4
FSL F4-0134	1.5	3.1	4	3.8	2.8	4.2	-1.9	-3.4
FSL F4-0152	4.9	4.6	4.6	5.1	4.8	4.9	-0.1	-2.1
FSL F4-0242	5.8	5.3	5.5	4.6	5.8	5.2	-2.1	-0.47
FSL H3-0280	4.1	2.7	3.9	3.3	4.7	3.4	3.1	4
FSL H3-0287	4.4	5.8	2.6	5.1	3.2	4.9	1.3	-0.04
FSL H3-0464	3.1	4.2	2	4	1.5	0.49	0.38	-0.13
FSL H3-0465	5.4	5.4	1.7	4.4	5.9	6	-0.79	-1.9
FSL H7-0433	3.1	4.7	2.7	4.7	3.5	5.2	-1.8	-2.6
FSL H7-0443	6	5.2	6.1	7.8	6	6	-0.54	1
FSL H7-0604	5.3	4.7	2.8	4.7	4.9	4.9	-1.4	-1.4
FSL H7-0694	5.8	5.2	5.7	6.2	6.1	6.2	0.35	2.4
FSL H7-0710	5.5	5.2	4.3	4.8	4.4	4.7	0.94	3.2
FSL H7-0713	7.5	7.4	5.8	4.8	6.1	6.2	-1	2.8
FSL H7-0718	4.8	4.7	5.5	5.8	3.7	5.5	-0.76	1.2
FSL H7-0918	4.6	4.8	4.7	5.5	5.3	5.7	-2	-2
FSL H8-0069	5.6	5.4	5.3	6.4	5.4	5.5	-1.7	0.24
FSL H8-0147	5.1	5.1	1.5	5.2	4.8	5	-2	-0.6
FSL H8-0175	3.8	3.3	3.6	4.2	3.6	3.6	0.44	2.7
FSL H8-0237	5	4.7	3.8	4	3.6	3.8	1.3	1.6
FSL J3-0153	4.3	4.3	4.3	3.9	-0.02	3.2	-1.5	1.9
FSL J3-0155	3.7	3.6	2.5	3.2	3.6	3.7	0.21	3.4
FSL J3-0159	4	3.5	4.7	4.5	3.9	3.9	3	3.8
FLS R5-0636	5.6	5.7	3.5	3.9	4.5	4.3	3.4	3.9

FSL R5-0883	6	6.1	2.7	4.9	5	5.5	1.4	2.9
FSL R5-0923	6	5.5	4.7	6.2	6	6.2	-0.46	0.53
FSL R5-0937	6.5	6.5	1.7	5.2	5.2	5.4	-1	5

*Counts are averaged for the three replicates and normalized to day 0 counts.

Supplemental Table 3.4: Hidden Markov Model (HMM) results for 25 *P. odorifer* and 3 closely related isolates

Query	Accession	Strain	CountMatchPerGenome[a]	CountGenomes[b]
Caps_synth_CapC	PF14102.5	FSL_F4-0134	1	1
Caps_synth_CapC	PF14102.5	FSL_H7-0433	1	2
Caps_synth_CapC	PF14102.5	FSL_H7-0443	2	3
Caps_synth_CapC	PF14102.5	FSL_H7-0710	2	4
Caps_synth_CapC	PF14102.5	FSL_H7-0713	1	5
Caps_synth_CapC	PF14102.5	FSL_H8-0175	1	6
Caps_synth_CapC	PF14102.5	FSL_J3-0155	1	7
Caps_synth_CapC	PF14102.5	FSL_R5-0923	2	8
CSD	PF00313.21	FSL_F4-0085	6	1
CSD	PF00313.21	FSL_F4-0126	6	2
CSD	PF00313.21	FSL_F4-0134	7	3
CSD	PF00313.21	FSL_F4-0152	6	4
CSD	PF00313.21	FSL_F4-0242	6	5
CSD	PF00313.21	FSL_H3-0280	6	6
CSD	PF00313.21	FSL_H3-0287	6	7
CSD	PF00313.21	FSL_H3-0464	6	8
CSD	PF00313.21	FSL_H3-0465	6	9
CSD	PF00313.21	FSL_H7-0433	6	10
CSD	PF00313.21	FSL_H7-0443	9	11
CSD	PF00313.21	FSL_H7-0604	6	12
CSD	PF00313.21	FSL_H7-0694	6	13
CSD	PF00313.21	FSL_H7-0710	9	14
CSD	PF00313.21	FSL_H7-0713	7	15
CSD	PF00313.21	FSL_H7-0718	7	16
CSD	PF00313.21	FSL_H7-0918	6	17
CSD	PF00313.21	FSL_H8-0069	7	18
CSD	PF00313.21	FSL_H8-0147	6	19
CSD	PF00313.21	FSL_H8-0175	6	20
CSD	PF00313.21	FSL_H8-0237	6	21
CSD	PF00313.21	FSL_J3-0153	6	22
CSD	PF00313.21	FSL_J3-0155	6	23
CSD	PF00313.21	FSL_J3-0159	6	24
CSD	PF00313.21	FSL_R5-0636	6	25
CSD	PF00313.21	FSL_R5-0883	6	26
CSD	PF00313.21	FSL_R5-0923	9	27
CSD	PF00313.21	FSL_R5-0937	6	28
DEAD	PF00270.28	FSL_F4-0085	40	1
DEAD	PF00270.28	FSL_F4-0126	42	2

DEAD	PF00270.28	FSL_F4-0134	46	3
DEAD	PF00270.28	FSL_F4-0152	45	4
DEAD	PF00270.28	FSL_F4-0242	43	5
DEAD	PF00270.28	FSL_H3-0280	42	6
DEAD	PF00270.28	FSL_H3-0287	43	7
DEAD	PF00270.28	FSL_H3-0464	49	8
DEAD	PF00270.28	FSL_H3-0465	45	9
DEAD	PF00270.28	FSL_H7-0433	45	10
DEAD	PF00270.28	FSL_H7-0443	40	11
DEAD	PF00270.28	FSL_H7-0604	47	12
DEAD	PF00270.28	FSL_H7-0694	41	13
DEAD	PF00270.28	FSL_H7-0710	49	14
DEAD	PF00270.28	FSL_H7-0713	42	15
DEAD	PF00270.28	FSL_H7-0718	43	16
DEAD	PF00270.28	FSL_H7-0918	48	17
DEAD	PF00270.28	FSL_H8-0069	41	18
DEAD	PF00270.28	FSL_H8-0147	43	19
DEAD	PF00270.28	FSL_H8-0175	43	20
DEAD	PF00270.28	FSL_H8-0237	46	21
DEAD	PF00270.28	FSL_J3-0153	45	22
DEAD	PF00270.28	FSL_J3-0155	42	23
DEAD	PF00270.28	FSL_J3-0159	41	24
DEAD	PF00270.28	FSL_R5-0636	42	25
DEAD	PF00270.28	FSL_R5-0883	44	26
DEAD	PF00270.28	FSL_R5-0923	40	27
DEAD	PF00270.28	FSL_R5-0937	45	28
DnaJ	PF00226.30	FSL_F4-0085	3	1
DnaJ	PF00226.30	FSL_F4-0126	3	2
DnaJ	PF00226.30	FSL_F4-0134	3	3
DnaJ	PF00226.30	FSL_F4-0152	3	4
DnaJ	PF00226.30	FSL_F4-0242	3	5
DnaJ	PF00226.30	FSL_H3-0280	3	6
DnaJ	PF00226.30	FSL_H3-0287	3	7
DnaJ	PF00226.30	FSL_H3-0464	3	8
DnaJ	PF00226.30	FSL_H3-0465	3	9
DnaJ	PF00226.30	FSL_H7-0433	3	10
DnaJ	PF00226.30	FSL_H7-0443	4	11
DnaJ	PF00226.30	FSL_H7-0604	3	12
DnaJ	PF00226.30	FSL_H7-0694	3	13
DnaJ	PF00226.30	FSL_H7-0710	3	14
DnaJ	PF00226.30	FSL_H7-0713	3	15
DnaJ	PF00226.30	FSL_H7-0718	4	16

DnaJ	PF00226.30	FSL_H7-0918	3	17
DnaJ	PF00226.30	FSL_H8-0069	3	18
DnaJ	PF00226.30	FSL_H8-0147	3	19
DnaJ	PF00226.30	FSL_H8-0175	3	20
DnaJ	PF00226.30	FSL_H8-0237	3	21
DnaJ	PF00226.30	FSL_J3-0153	3	22
DnaJ	PF00226.30	FSL_J3-0155	3	23
DnaJ	PF00226.30	FSL_J3-0159	3	24
DnaJ	PF00226.30	FSL_R5-0636	3	25
DnaJ	PF00226.30	FSL_R5-0883	3	26
DnaJ	PF00226.30	FSL_R5-0923	3	27
DnaJ	PF00226.30	FSL_R5-0937	3	28
FA_desaturase	PF00487.23	FSL_F4-0085	5	1
FA_desaturase	PF00487.23	FSL_F4-0126	5	2
FA_desaturase	PF00487.23	FSL_F4-0134	5	3
FA_desaturase	PF00487.23	FSL_F4-0152	5	4
FA_desaturase	PF00487.23	FSL_F4-0242	5	5
FA_desaturase	PF00487.23	FSL_H3-0280	5	6
FA_desaturase	PF00487.23	FSL_H3-0287	6	7
FA_desaturase	PF00487.23	FSL_H3-0464	5	8
FA_desaturase	PF00487.23	FSL_H3-0465	4	9
FA_desaturase	PF00487.23	FSL_H7-0433	5	10
FA_desaturase	PF00487.23	FSL_H7-0443	5	11
FA_desaturase	PF00487.23	FSL_H7-0604	5	12
FA_desaturase	PF00487.23	FSL_H7-0694	5	13
FA_desaturase	PF00487.23	FSL_H7-0710	7	14
FA_desaturase	PF00487.23	FSL_H7-0713	5	15
FA_desaturase	PF00487.23	FSL_H7-0718	5	16
FA_desaturase	PF00487.23	FSL_H7-0918	5	17
FA_desaturase	PF00487.23	FSL_H8-0069	4	18
FA_desaturase	PF00487.23	FSL_H8-0147	5	19
FA_desaturase	PF00487.23	FSL_H8-0175	4	20
FA_desaturase	PF00487.23	FSL_H8-0237	6	21
FA_desaturase	PF00487.23	FSL_J3-0153	5	22
FA_desaturase	PF00487.23	FSL_J3-0155	5	23
FA_desaturase	PF00487.23	FSL_J3-0159	5	24
FA_desaturase	PF00487.23	FSL_R5-0636	4	25
FA_desaturase	PF00487.23	FSL_R5-0883	5	26
FA_desaturase	PF00487.23	FSL_R5-0923	5	27
FA_desaturase	PF00487.23	FSL_R5-0937	3	28
FA_desaturase_2	PF03405.13	FSL_F4-0085	2	1
FA_desaturase_2	PF03405.13	FSL_F4-0126	2	2

FA_desaturase_2	PF03405.13	FSL_F4-0134	2	3
FA_desaturase_2	PF03405.13	FSL_F4-0152	2	4
FA_desaturase_2	PF03405.13	FSL_F4-0242	2	5
FA_desaturase_2	PF03405.13	FSL_H3-0280	2	6
FA_desaturase_2	PF03405.13	FSL_H3-0287	2	7
FA_desaturase_2	PF03405.13	FSL_H3-0464	2	8
FA_desaturase_2	PF03405.13	FSL_H3-0465	2	9
FA_desaturase_2	PF03405.13	FSL_H7-0433	2	10
FA_desaturase_2	PF03405.13	FSL_H7-0443	1	11
FA_desaturase_2	PF03405.13	FSL_H7-0604	2	12
FA_desaturase_2	PF03405.13	FSL_H7-0694	3	13
FA_desaturase_2	PF03405.13	FSL_H7-0710	1	14
FA_desaturase_2	PF03405.13	FSL_H7-0713	2	15
FA_desaturase_2	PF03405.13	FSL_H7-0718	2	16
FA_desaturase_2	PF03405.13	FSL_H7-0918	2	17
FA_desaturase_2	PF03405.13	FSL_H8-0069	2	18
FA_desaturase_2	PF03405.13	FSL_H8-0147	2	19
FA_desaturase_2	PF03405.13	FSL_H8-0175	2	20
FA_desaturase_2	PF03405.13	FSL_H8-0237	2	21
FA_desaturase_2	PF03405.13	FSL_J3-0153	2	22
FA_desaturase_2	PF03405.13	FSL_J3-0155	2	23
FA_desaturase_2	PF03405.13	FSL_J3-0159	2	24
FA_desaturase_2	PF03405.13	FSL_R5-0636	2	25
FA_desaturase_2	PF03405.13	FSL_R5-0883	2	26
FA_desaturase_2	PF03405.13	FSL_R5-0923	1	27
FA_desaturase_2	PF03405.13	FSL_R5-0937	2	28
LtrA	PF06772.10	FSL_F4-0085	2	1
LtrA	PF06772.10	FSL_F4-0126	2	2
LtrA	PF06772.10	FSL_F4-0134	2	3
LtrA	PF06772.10	FSL_F4-0152	2	4
LtrA	PF06772.10	FSL_F4-0242	2	5
LtrA	PF06772.10	FSL_H3-0280	2	6
LtrA	PF06772.10	FSL_H3-0287	2	7
LtrA	PF06772.10	FSL_H3-0464	2	8
LtrA	PF06772.10	FSL_H3-0465	2	9
LtrA	PF06772.10	FSL_H7-0433	1	10
LtrA	PF06772.10	FSL_H7-0443	1	11
LtrA	PF06772.10	FSL_H7-0604	2	12
LtrA	PF06772.10	FSL_H7-0694	2	13
LtrA	PF06772.10	FSL_H7-0710	1	14
LtrA	PF06772.10	FSL_H7-0713	2	15
LtrA	PF06772.10	FSL_H7-0718	2	16

LtrA	PF06772.10	FSL_H7-0918	2	17
LtrA	PF06772.10	FSL_H8-0069	2	18
LtrA	PF06772.10	FSL_H8-0147	2	19
LtrA	PF06772.10	FSL_H8-0175	2	20
LtrA	PF06772.10	FSL_H8-0237	2	21
LtrA	PF06772.10	FSL_J3-0153	2	22
LtrA	PF06772.10	FSL_J3-0155	2	23
LtrA	PF06772.10	FSL_J3-0159	2	24
LtrA	PF06772.10	FSL_R5-0636	2	25
LtrA	PF06772.10	FSL_R5-0883	2	26
LtrA	PF06772.10	FSL_R5-0923	1	27
LtrA	PF06772.10	FSL_R5-0937	2	28
Peptidase_S11	PF00768.19	FSL_F4-0085	7	1
Peptidase_S11	PF00768.19	FSL_F4-0126	9	2
Peptidase_S11	PF00768.19	FSL_F4-0134	9	3
Peptidase_S11	PF00768.19	FSL_F4-0152	7	4
Peptidase_S11	PF00768.19	FSL_F4-0242	7	5
Peptidase_S11	PF00768.19	FSL_H3-0280	7	6
Peptidase_S11	PF00768.19	FSL_H3-0287	7	7
Peptidase_S11	PF00768.19	FSL_H3-0464	8	8
Peptidase_S11	PF00768.19	FSL_H3-0465	8	9
Peptidase_S11	PF00768.19	FSL_H7-0433	8	10
Peptidase_S11	PF00768.19	FSL_H7-0443	9	11
Peptidase_S11	PF00768.19	FSL_H7-0604	9	12
Peptidase_S11	PF00768.19	FSL_H7-0694	7	13
Peptidase_S11	PF00768.19	FSL_H7-0710	6	14
Peptidase_S11	PF00768.19	FSL_H7-0713	7	15
Peptidase_S11	PF00768.19	FSL_H7-0718	8	16
Peptidase_S11	PF00768.19	FSL_H7-0918	7	17
Peptidase_S11	PF00768.19	FSL_H8-0069	7	18
Peptidase_S11	PF00768.19	FSL_H8-0147	8	19
Peptidase_S11	PF00768.19	FSL_H8-0175	9	20
Peptidase_S11	PF00768.19	FSL_H8-0237	9	21
Peptidase_S11	PF00768.19	FSL_J3-0153	8	22
Peptidase_S11	PF00768.19	FSL_J3-0155	9	23
Peptidase_S11	PF00768.19	FSL_J3-0159	8	24
Peptidase_S11	PF00768.19	FSL_R5-0636	8	25
Peptidase_S11	PF00768.19	FSL_R5-0883	9	26
Peptidase_S11	PF00768.19	FSL_R5-0923	9	27
Peptidase_S11	PF00768.19	FSL_R5-0937	9	28
RecA	PF00154.20	FSL_F4-0085	2	1
RecA	PF00154.20	FSL_F4-0126	3	2

RecA	PF00154.20	FSL_F4-0134	2	3
RecA	PF00154.20	FSL_F4-0152	3	4
RecA	PF00154.20	FSL_F4-0242	3	5
RecA	PF00154.20	FSL_H3-0280	3	6
RecA	PF00154.20	FSL_H3-0287	2	7
RecA	PF00154.20	FSL_H3-0464	3	8
RecA	PF00154.20	FSL_H3-0465	2	9
RecA	PF00154.20	FSL_H7-0433	3	10
RecA	PF00154.20	FSL_H7-0443	4	11
RecA	PF00154.20	FSL_H7-0604	3	12
RecA	PF00154.20	FSL_H7-0694	3	13
RecA	PF00154.20	FSL_H7-0710	3	14
RecA	PF00154.20	FSL_H7-0713	2	15
RecA	PF00154.20	FSL_H7-0718	3	16
RecA	PF00154.20	FSL_H7-0918	3	17
RecA	PF00154.20	FSL_H8-0069	3	18
RecA	PF00154.20	FSL_H8-0147	3	19
RecA	PF00154.20	FSL_H8-0175	3	20
RecA	PF00154.20	FSL_H8-0237	3	21
RecA	PF00154.20	FSL_J3-0153	3	22
RecA	PF00154.20	FSL_J3-0155	3	23
RecA	PF00154.20	FSL_J3-0159	3	24
RecA	PF00154.20	FSL_R5-0636	3	25
RecA	PF00154.20	FSL_R5-0883	3	26
RecA	PF00154.20	FSL_R5-0923	3	27
RecA	PF00154.20	FSL_R5-0937	3	28
YdjO	PF14169.5	FSL_F4-0085	3	1
YdjO	PF14169.5	FSL_F4-0126	3	2
YdjO	PF14169.5	FSL_F4-0134	3	3
YdjO	PF14169.5	FSL_F4-0152	3	4
YdjO	PF14169.5	FSL_F4-0242	3	5
YdjO	PF14169.5	FSL_H3-0280	3	6
YdjO	PF14169.5	FSL_H3-0287	3	7
YdjO	PF14169.5	FSL_H3-0464	3	8
YdjO	PF14169.5	FSL_H3-0465	4	9
YdjO	PF14169.5	FSL_H7-0433	3	10
YdjO	PF14169.5	FSL_H7-0443	3	11
YdjO	PF14169.5	FSL_H7-0604	3	12
YdjO	PF14169.5	FSL_H7-0694	3	13
YdjO	PF14169.5	FSL_H7-0710	3	14
YdjO	PF14169.5	FSL_H7-0713	3	15
YdjO	PF14169.5	FSL_H7-0718	3	16

YdjO	PF14169.5	FSL_H7-0918	3	17
YdjO	PF14169.5	FSL_H8-0069	3	18
YdjO	PF14169.5	FSL_H8-0147	3	19
YdjO	PF14169.5	FSL_H8-0175	3	20
YdjO	PF14169.5	FSL_H8-0237	3	21
YdjO	PF14169.5	FSL_J3-0153	3	22
YdjO	PF14169.5	FSL_J3-0155	3	23
YdjO	PF14169.5	FSL_J3-0159	3	24
YdjO	PF14169.5	FSL_R5-0636	3	25
YdjO	PF14169.5	FSL_R5-0883	3	26
YdjO	PF14169.5	FSL_R5-0923	3	27
YdjO	PF14169.5	FSL_R5-0937	3	28

[a]CountMatchPerGenome: The number of times a HMM Query was found in a given genome
[b]CountGenomes: A running total of the number of genomes in which a HMM Query was found

Supplemental Table 3.5: List of genes enriched in phylogenetic clade A

Cluster	Number of Genes	Number of Taxa	Clade A Presence	Clade B Presence	Clade A Absence	Clade B Absence	P-value [a]	FDR P-value [b]	Odds Ratio [c]	Products
Cluster_5417	15	15	13	2	0	10	0.0000202	0.001189034	INF	ABC transporter ATP-binding protein
Cluster_5735	10	10	10	0	3	12	0.000108	0.005315681	INF	ABC transporter ATP-binding protein
Cluster_5569	13	13	13	0	0	12	0.000000192	0.0000258	INF	ABC transporter permease
Cluster_5544	13	13	13	0	0	12	0.000000192	0.0000258	INF	acetolactate synthase, large subunit,
Cluster_5385	16	16	12	2	1	10	0.00021268	0.0101023	45.039313 77	acetylhydrolase
Cluster_5562	13	13	13	0	0	12	0.000000192	0.0000258	INF	anthranilate phosphoribosyltransferase
Cluster_5527	13	13	12	0	1	12	0.00000269	0.000213464	INF	CAAX protease family protein
Cluster_5682	11	11	11	0	2	12	0.0000202	0.001189034	INF	cation-binding protein
Cluster_5383	16	16	13	2	0	10	0.0000202	0.001189034	INF	chemotaxis protein
Cluster_5539	13	13	13	0	0	12	0.000000192	0.0000258	INF	cyclic pyranopterin phosphate synthase
Cluster_5490	14	14	13	1	0	11	0.00000269	0.000213464	INF	DNA mismatch repair protein MutT
Cluster_5536	13	13	13	0	0	12	0.000000192	0.0000258	INF	DNA-binding response regulator
Cluster_5503	14	14	12	2	1	10	0.00021268	0.0101023	45.039313 77	DNA-binding response regulator
Cluster_5619	12	12	11	1	2	11	0.00021268	0.0101023	45.405431 87	DoxX family protein
Cluster_5318	18	18	12	3	1	9	0.000982636	0.03850249	29.256081 75	enamine deaminase RidA
Cluster_5558	13	13	13	0	0	12	0.000000192	0.0000258	INF	formate transporter
Cluster_5534	13	13	13	0	0	12	0.000000192	0.0000258	INF	GAF domain-containing protein
Cluster_5543	13	13	12	1	1	11	0.0000327	0.001859845	82.675413	GNAT family N-acetyltransferase
Cluster_5394	16	16	13	3	0	9	0.000107686	0.005315681	INF	GNAT family N-acetyltransferase
Cluster_5337	18	18	12	3	1	9	0.000982636	0.03850249	29.256081 75	GNAT family N-acetyltransferase
Cluster_5608	12	12	12	0	1	12	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5873	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein

Cluster_5528	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5529	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5790	9	9	9	0	4	12	0.000457666	0.01961147	INF	hypothetical protein
Cluster_5540	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5713	10	10	10	0	3	12	0.000108	0.005315681	INF	hypothetical protein
Cluster_5416	15	15	13	2	0	10	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5382	16	16	13	3	0	9	0.000107686	0.005315681	INF	hypothetical protein
Cluster_5791	9	9	9	0	4	12	0.000457666	0.01961147	INF	hypothetical protein
Cluster_5792	9	9	9	0	4	12	0.000457666	0.01961147	INF	hypothetical protein
Cluster_5715	10	10	9	0	4	12	0.000457666	0.01961147	INF	hypothetical protein
Cluster_5418	15	15	13	2	0	10	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5483	14	14	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5611	12	12	12	0	1	12	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5419	15	15	12	3	1	9	0.000982636	0.03850249	29.256081 75	hypothetical protein
Cluster_5420	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5421	15	15	13	2	0	10	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5545	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5546	13	13	12	1	1	11	0.0000327	0.001859845	82.675413	hypothetical protein
Cluster_5876	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5683	11	11	11	0	2	12	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5342	17	17	13	3	0	9	0.000107686	0.005315681	INF	hypothetical protein
Cluster_5320	18	17	13	3	0	9	0.000107686	0.005315681	INF	hypothetical protein
Cluster_5343	17	17	13	3	0	9	0.000107686	0.005315681	INF	hypothetical protein
Cluster_5548	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5686	11	11	11	0	2	12	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5549	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5550	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5551	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5552	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5617	12	12	12	0	1	12	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5690	11	11	11	0	2	12	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5553	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5323	18	18	13	5	0	7	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5555	13	13	11	2	2	10	0.001202623	0.0460472	22.436347 12	hypothetical protein

Cluster_5425	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5492	14	14	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5426	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5493	14	14	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5427	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5428	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5429	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5494	14	14	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5431	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5495	14	14	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5432	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5496	14	14	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5433	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5434	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5436	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5437	15	15	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5497	14	14	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5884	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5818	9	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5886	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5887	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5889	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5325	18	18	13	5	0	7	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5498	14	14	12	0	1	12	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5556	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5500	14	14	12	1	1	11	0.0000327	0.001859845	82.675413	hypothetical protein
Cluster_5561	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5501	14	14	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5563	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5389	16	16	13	3	0	9	0.000107686	0.005315681	INF	hypothetical protein
Cluster_5564	13	13	10	1	3	11	0.000982636	0.03850249	29.782730 38	hypothetical protein
Cluster_5728	10	10	10	0	3	12	0.000108	0.005315681	INF	hypothetical protein
Cluster_5391	16	16	13	3	0	9	0.000107686	0.005315681	INF	hypothetical protein
Cluster_5639	12	12	12	0	1	12	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5570	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein

Cluster_5392	16	16	13	3	0	9	0.000107686	0.005315681	INF	hypothetical protein
Cluster_5362	17	17	13	4	0	8	0.000457666	0.01961147	INF	hypothetical protein
Cluster_5363	17	17	13	4	0	8	0.000457666	0.01961147	INF	hypothetical protein
Cluster_5502	14	14	12	2	1	10	0.00021268	0.0101023	45.039313 77	hypothetical protein
Cluster_5730	10	10	10	0	3	12	0.000108	0.005315681	INF	hypothetical protein
Cluster_5731	10	10	10	0	3	12	0.000108	0.005315681	INF	hypothetical protein
Cluster_5732	10	10	10	0	3	12	0.000108	0.005315681	INF	hypothetical protein
Cluster_5733	10	10	10	0	3	12	0.000108	0.005315681	INF	hypothetical protein
Cluster_5736	10	10	10	0	3	12	0.000108	0.005315681	INF	hypothetical protein
Cluster_5329	18	18	13	5	0	7	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5641	12	12	12	0	1	12	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5642	12	12	12	0	1	12	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5897	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5643	12	12	12	0	1	12	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5504	14	14	12	2	1	10	0.00021268	0.0101023	45.039313 77	hypothetical protein
Cluster_5505	14	14	12	2	1	10	0.00021268	0.0101023	45.039313 77	hypothetical protein
Cluster_5830	9	9	9	0	4	12	0.000457666	0.01961147	INF	hypothetical protein
Cluster_5573	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5574	13	13	13	0	0	12	0.000000192	0.0000258	INF	hypothetical protein
Cluster_5739	10	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5506	14	14	13	1	0	11	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5441	15	15	13	2	0	10	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5333	18	18	13	2	0	10	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5647	12	12	12	0	1	12	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5648	12	12	12	0	1	12	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5442	15	15	13	2	0	10	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5749	10	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5750	10	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5751	10	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5700	11	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5336	18	18	13	5	0	7	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5906	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5701	11	11	11	0	2	12	0.0000202	0.001189034	INF	hypothetical protein

Cluster_5702	11	11	11	0	2	12	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5703	11	11	11	0	2	12	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5451	15	15	12	0	1	12	0.00000269	0.000213464	INF	hypothetical protein
Cluster_5705	11	11	11	0	2	12	0.0000202	0.001189034	INF	hypothetical protein
Cluster_5836	9	9	9	0	4	12	0.000457666	0.01961147	INF	hypothetical protein
Cluster_5918	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5919	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5933	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5777	10	10	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5935	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5936	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5937	8	8	8	0	5	12	0.001647597	0.04937888	INF	hypothetical protein
Cluster_5938	8	8	8	0	5	12	0.001647597	0.04937888	INF	IS110 family transposase
Cluster_5861	9	9	8	0	5	12	0.001647597	0.04937888	INF	IS110 family transposase
Cluster_5485	14	14	12	2	1	10	0.00021268	0.0101023	45.039313 77	MerR family transcriptional regulator
Cluster_5685	11	11	11	0	2	12	0.0000202	0.001189034	INF	MerR family transcriptional regulator
Cluster_5344	17	17	13	2	0	10	0.0000202	0.001189034	INF	MerR family transcriptional regulator
Cluster_5734	10	10	10	0	3	12	0.000108	0.005315681	INF	methionine ABC transporter permease
Cluster_5566	13	13	13	0	0	12	0.000000192	0.0000258	INF	MFS transporter
Cluster_5888	8	8	8	0	5	12	0.001647597	0.04937888	INF	minor capsid protein
Cluster_5542	13	13	13	0	0	12	0.000000192	0.0000258	INF	molybdenum cofactor biosynthesis protein MoaE
Cluster_5538	13	13	13	0	0	12	0.000000192	0.0000258	INF	molybdopterin molybdenumtransferase
Cluster_5541	13	13	13	0	0	12	0.000000192	0.0000258	INF	molybdopterin-guanine dinucleotide biosynthesis
Cluster_5575	13	13	13	0	0	12	0.000000192	0.0000258	INF	N-acetyltransferase
Cluster_5554	13	13	11	2	2	10	0.001202623	0.0460472	22.436347 12	NAD(P)-dependent oxidoreductase
Cluster_5532	13	13	13	0	0	12	0.000000192	0.0000258	INF	nitrate reductase molybdenum cofactor assembly
Cluster_5530	13	13	13	0	0	12	0.000000192	0.0000258	INF	nitrate reductase subunit alpha
Cluster_5531	13	13	13	0	0	12	0.000000192	0.0000258	INF	nitrate reductase subunit beta

Cluster_5537	13	13	13	0	0	12	0.000000192	0.0000258	INF	nitrate/nitrite transporter
Cluster_5559	13	13	13	0	0	12	0.000000192	0.0000258	INF	nitrite reductase
Cluster_5560	13	13	13	0	0	12	0.000000192	0.0000258	INF	nitrite reductase large subunit
Cluster_5704	11	11	10	1	3	11	0.000982636	0.03850249	29.782730 38	O-acetyl-ADP-ribose deacetylase
Cluster_5567	13	13	13	0	0	12	0.000000192	0.0000258	INF	phosphoesterase
Cluster_5414	15	15	13	2	0	10	0.0000202	0.001189034	INF	radical SAM/SPASM domain-containing protein
Cluster_5415	15	15	13	2	0	10	0.0000202	0.001189034	INF	radical SAM/SPASM domain-containing protein
Cluster_5533	13	13	13	0	0	12	0.000000192	0.0000258	INF	respiratory nitrate reductase subunit gamma
Cluster_5499	14	14	12	1	1	11	0.0000327	0.001859845	82.675413	RNA polymerase
Cluster_5393	16	16	13	3	0	9	0.000107686	0.005315681	INF	RNA polymerase subunit sigma
Cluster_5571	13	13	11	2	2	10	0.001202623	0.0460472	22.436347 12	sensor histidine kinase
Cluster_5568	13	13	13	0	0	12	0.000000192	0.0000258	INF	sugar ABC transporter ATP-binding protein
Cluster_5565	13	13	13	0	0	12	0.000000192	0.0000258	INF	thioredoxin
Cluster_5526	13	13	12	0	1	12	0.00000269	0.000213464	INF	transcriptional regulator
Cluster_5405	16	10	8	0	5	12	0.001647597	0.04937888	INF	transposase
Cluster_5535	13	13	13	0	0	12	0.000000192	0.0000258	INF	two-component sensor histidine kinase
Cluster_5435	15	15	13	1	0	11	0.00000269	0.000213464	INF	type II secretion system protein E
Cluster_5430	15	15	13	1	0	11	0.00000269	0.000213464	INF	type VII secretion protein EssC
Cluster_5557	13	13	13	0	0	12	0.000000192	0.0000258	INF	uroporphyrinogen-III C-methyltransferase
Cluster_5684	11	11	11	0	2	12	0.0000202	0.001189034	INF	Zn-dependent protease

[a] P-values from two-sided Fisher's Exact Tests

[b] P-values were corrected using the False Discovery Rate (FDR) approach

[c] Odds ratios marked as INF (Infinite) are a result of dividing by zero

Supplemental Table 3.6: List of genes enriched in phylogenetic clade B

Cluster	Number of Genes	Number of Taxa	Clade A Presence	Clade B Presence	Clade A Absence	Clade B Absence	P-value [a]	FDR P-value [b]	Odds Ratio	Products
Cluster_5352	17	17	5	12	8	0	0.001647597	0.04937888	0	3D-(3,5/4)-trihydroxycyclohexane-1,2-dione
Cluster_5593	13	13	1	12	12	0	0.00000269	0.000213464	0	4-carboxymuconolactone decarboxylase
Cluster_5350	17	17	5	12	8	0	0.001647597	0.04937888	0	5-dehydro-2-deoxygluconokinase
Cluster_5348	17	17	5	12	8	0	0.001647597	0.04937888	0	5-deoxy-glucuronate isomerase
Cluster_5282	20	20	5	12	8	0	0.001647597	0.04937888	0	ABC transporter
Cluster_5782	10	10	0	10	13	2	0.0000202	0.001189034	0	ABC transporter ATP-binding protein
Cluster_5786	10	10	0	10	13	2	0.0000202	0.001189034	0	ABC transporter ATP-binding protein
Cluster_5357	17	17	5	12	8	0	0.001647597	0.04937888	0	ABC transporter permease
Cluster_5670	12	12	0	12	13	0	0.000000192	0.0000258	0	ABC transporter permease
Cluster_5467	15	15	0	12	13	0	0.000000192	0.0000258	0	ABC transporter permease
Cluster_5477	15	15	0	12	13	0	0.000000192	0.0000258	0	ABC transporter permease
Cluster_5478	15	15	0	12	13	0	0.000000192	0.0000258	0	ABC transporter permease
Cluster_6084	7	7	0	7	13	5	0.001647597	0.04937888	0	ABC transporter permease
Cluster_5785	10	10	0	10	13	2	0.0000202	0.001189034	0	ABC transporter permease
Cluster_5355	17	17	5	12	8	0	0.001647597	0.04937888	0	ABC transporter substrate-binding protein
Cluster_5667	12	12	2	10	11	2	0.001202623	0.0460472	0.0445705 35	ABC transporter substrate-binding protein
Cluster_5479	15	15	0	12	13	0	0.000000192	0.0000258	0	ABC transporter substrate-binding protein
Cluster_5679	12	12	0	10	13	2	0.0000202	0.001189034	0	ABC transporter substrate-binding protein
Cluster_6094	7	7	0	7	13	5	0.001647597	0.04937888	0	ABC transporter substrate-binding protein
Cluster_5867	9	9	0	8	13	4	0.000457666	0.01961147	0	alcohol dehydrogenase
Cluster_5712	11	11	0	8	13	4	0.000457666	0.01961147	0.0341809 0	alpha-amylase
Cluster_5658	12	12	1	9	12	3	0.000982636	0.03850249	27	alpha/beta hydrolase
Cluster_5523	14	14	0	11	13	1	0.00000269	0.000213464	0	antitoxin
Cluster_5462	15	15	0	12	13	0	0.000000192	0.0000258	0	arabinose ABC transporter permease
Cluster_5345	17	17	5	12	8	0	0.001647597	0.04937888	0	AraC family transcriptional regulator
Cluster_5354	17	17	5	12	8	0	0.001647597	0.04937888	0	AraC family transcriptional regulator
Cluster_5708	11	11	0	11	13	1	0.00000269	0.000213464	0	AraC family transcriptional regulator

Cluster_5669	12	12	0	12	13	0	0.000000192	0.0000258	0	AraC family transcriptional regulator
Cluster_5604	13	13	0	10	13	2	0.0000202	0.001189034	0	AraC family transcriptional regulator
Cluster_5966	8	8	0	8	13	4	0.000457666	0.01961147	0	AsnC family transcriptional regulator
Cluster_5461	15	15	0	12	13	0	0.000000192	0.0000258	0	cysteine desulfurase NifS
Cluster_5410	16	16	1	12	12	0	0.00000269	0.000213464	0	deoxyribonuclease IV
Cluster_5595	13	13	1	12	12	0	0.00000269	0.000213464	0	dihydrolipoyl dehydrogenase
Cluster_5599	13	13	0	10	13	2	0.0000202	0.001189034	0	dihydrolipoyl dehydrogenase
Cluster_5780	10	10	0	10	13	2	0.0000202	0.001189034	0	DNA-binding response regulator
Cluster_6081	7	7	0	7	13	5	0.001647597	0.04937888	0	DNA-binding response regulator
Cluster_6091	7	7	0	7	13	5	0.001647597	0.04937888	0	DNA-binding response regulator
Cluster_5277	20	20	5	12	8	0	0.001647597	0.04937888	0	flotillin
Cluster_5680	12	12	0	11	13	1	0.00000269	0.000213464	0	FMN reductase (NADPH)
Cluster_5675	12	12	0	10	13	2	0.0000202	0.001189034	0	fumarylacetoacetate hydrolase
Cluster_5423	15	15	3	12	10	0	0.000107686	0.005315681	0	glycoside hydrolase
Cluster_5768	10	10	1	9	12	3	0.000982636	0.03850249	0.034 1809 27	glycoside hydrolase
Cluster_5346	17	17	5	12	8	0	0.001647597	0.04937888	0	glycoside hydrolase 105 family protein
Cluster_5606	13	13	0	10	13	2	0.0000202	0.001189034	0	glycosyl hydrolase
Cluster_5455	15	15	1	12	12	0	0.00000269	0.000213464	0	GNAT family N-acetyltransferase
Cluster_5519	14	14	3	11	10	1	0.000982636	0.03850249	0.033 5765 05	GNAT family N-acetyltransferase
Cluster_5482	15	15	1	11	12	1	0.0000327	0.001859845	0.012 0954 94	GNAT family N-acetyltransferase
Cluster_5341	17	17	2	12	11	0	0.0000202	0.001189034	0	hypothetical protein
Cluster_5144	23	20	5	12	8	0	0.001647597	0.04937888	0	hypothetical protein
Cluster_5359	17	17	5	12	8	0	0.001647597	0.04937888	0	hypothetical protein
Cluster_5360	17	17	5	12	8	0	0.001647597	0.04937888	0	hypothetical protein
Cluster_5235	20	20	5	12	8	0	0.001647597	0.04937888	0	hypothetical protein
Cluster_5388	16	16	1	12	12	0	0.00000269	0.000213464	0	hypothetical protein
Cluster_5507	14	14	3	11	10	1	0.000982636	0.03850249	0.033 5765 05	hypothetical protein
Cluster_5398	16	16	4	12	9	0	0.000457666	0.01961147	0	hypothetical protein

Cluster_5401	16	16	4	12	9	0	0.000457666	0.01961147	0	hypothetical protein
Cluster_5306	19	17	4	12	9	0	0.000457666	0.01961147	0	hypothetical protein
Cluster_5402	16	16	1	12	12	0	0.00000269	0.000213464	0	hypothetical protein
Cluster_5767	10	10	1	9	12	3	0.000982636	0.03850249	0.034 1809 27	hypothetical protein
Cluster_5771	10	10	1	9	12	3	0.000982636	0.03850249	0.034 1809 27	hypothetical protein
Cluster_5772	10	10	1	9	12	3	0.000982636	0.03850249	0.034 1809 27	hypothetical protein
Cluster_5773	10	10	1	9	12	3	0.000982636	0.03850249	0.034 1809 27	hypothetical protein
Cluster_5375	17	17	2	12	11	0	0.0000202	0.001189034	0	hypothetical protein
Cluster_5776	10	10	1	9	12	3	0.000982636	0.03850249	0.034 1809 27	hypothetical protein
Cluster_5586	13	13	1	9	12	3	0.000982636	0.03850249	0.034 1809 27	hypothetical protein
Cluster_5778	10	10	1	9	12	3	0.000982636	0.03850249	0.034 1809 27	hypothetical protein
Cluster_5406	16	16	4	12	9	0	0.000457666	0.01961147	0	hypothetical protein
Cluster_5458	15	14	3	11	10	1	0.000982636	0.03850249	0.033 5765 05	hypothetical protein
Cluster_5592	13	13	2	11	11	1	0.00021268	0.0101023	0.022 0237 97	hypothetical protein
Cluster_5409	16	16	1	12	12	0	0.00000269	0.000213464	0	hypothetical protein
Cluster_5411	16	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5522	14	12	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5671	12	12	1	9	12	3	0.000982636	0.03850249	0.034 1809	hypothetical protein

									27	
Cluster_5463	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5863	9	9	0	9	13	3	0.000107686	0.005315681	0	hypothetical protein
Cluster_5597	13	13	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_6072	7	7	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
Cluster_5524	14	14	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5672	12	12	0	9	13	3	0.000107686	0.005315681	0	hypothetical protein
Cluster_5673	12	12	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5598	13	9	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
Cluster_5465	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5466	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5953	8	8	0	8	13	4	0.000457666	0.01961147	0	hypothetical protein
Cluster_5468	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5469	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5470	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5471	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5473	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5474	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5412	16	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5475	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5476	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5480	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5481	15	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_5413	16	15	0	12	13	0	0.000000192	0.0000258	0	hypothetical protein
Cluster_6074	7	7	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
Cluster_6075	7	7	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
Cluster_6076	7	7	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
Cluster_6077	7	7	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
Cluster_6078	7	7	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
Cluster_5866	9	9	0	9	13	3	0.000107686	0.005315681	0	hypothetical protein
Cluster_5781	10	10	0	10	13	2	0.0000202	0.001189034	0	hypothetical protein
Cluster_5957	8	8	0	8	13	4	0.000457666	0.01961147	0	hypothetical protein
Cluster_5958	8	8	0	8	13	4	0.000457666	0.01961147	0	hypothetical protein
Cluster_6085	7	7	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
Cluster_5868	9	9	0	9	13	3	0.000107686	0.005315681	0	hypothetical protein
Cluster_5869	9	9	0	9	13	3	0.000107686	0.005315681	0	hypothetical protein

Cluster_5870	9	9	0	9	13	3	0.000107686	0.005315681	0	hypothetical protein
Cluster_5600	13	13	0	10	13	2	0.0000202	0.001189034	0	hypothetical protein
Cluster_5677	12	12	0	10	13	2	0.0000202	0.001189034	0	hypothetical protein
Cluster_6086	7	7	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
Cluster_6087	7	7	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
Cluster_5961	8	8	0	8	13	4	0.000457666	0.01961147	0	hypothetical protein
Cluster_5678	12	12	0	10	13	2	0.0000202	0.001189034	0	hypothetical protein
Cluster_5871	9	9	0	9	13	3	0.000107686	0.005315681	0	hypothetical protein
Cluster_5962	8	8	0	8	13	4	0.000457666	0.01961147	0	hypothetical protein
Cluster_5605	13	13	0	10	13	2	0.0000202	0.001189034	0	hypothetical protein
Cluster_6088	7	7	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
									0.034 1809	
Cluster_5787	10	10	1	9	12	3	0.000982636	0.03850249	27	hypothetical protein
Cluster_5967	8	8	0	8	13	4	0.000457666	0.01961147	0	hypothetical protein
Cluster_5788	10	10	0	10	13	2	0.0000202	0.001189034	0	hypothetical protein
Cluster_6090	7	7	0	7	13	5	0.001647597	0.04937888	0	hypothetical protein
Cluster_5347	17	17	5	12	8	0	0.001647597	0.04937888	0	inositol 2-dehydrogenase
Cluster_5603	13	13	0	10	13	2	0.0000202	0.001189034	0	iron ABC transporter
Cluster_5602	13	13	0	10	13	2	0.0000202	0.001189034	0	iron ABC transporter permease
Cluster_5281	20	20	5	12	8	0	0.001647597	0.04937888	0	iron export ABC transporter permease subunit
Cluster_5349	17	17	5	12	8	0	0.001647597	0.04937888	0	LacI family transcriptional regulator
Cluster_5601	13	13	0	11	13	1	0.00000269	0.000213464	0	luciferase
Cluster_5674	12	12	0	12	13	0	0.000000192	0.0000258	0	lysine transporter LysE
Cluster_5710	11	11	0	11	13	1	0.00000269	0.000213464	0	LysR family transcriptional regulator
Cluster_5525	14	14	0	12	13	0	0.000000192	0.0000258	0	macrolide ABC transporter ATP-binding protein
Cluster_5594	13	13	1	12	12	0	0.00000269	0.000213464	0	MarR family transcriptional regulator
Cluster_5711	11	11	0	10	13	2	0.0000202	0.001189034	0	membrane-bound O-acyltransferase family protein
Cluster_5351	17	17	5	12	8	0	0.001647597	0.04937888	0	methylmalonate-semialdehyde dehydrogenase
Cluster_5965	8	8	0	8	13	4	0.000457666	0.01961147	0	MFS transporter
Cluster_5353	17	17	5	12	8	0	0.001647597	0.04937888	0	myo-inosose-2 dehydratase
Cluster_5516	14	14	1	11	12	1	0.0000327	0.001859845	0.012	NADPH:quinone reductase

								0.095494	
Cluster_5668	12	12	2	10	11	2	0.001202623	0.0460472	0.044570535	nitrilotriacetate monooxygenase
Cluster_6083	7	7	0	7	13	5	0.001647597	0.04937888	0	peptide ABC transporter ATP-binding protein
Cluster_5596	13	13	1	12	12	0	0.00000269	0.000213464	0	peroxiredoxin
Cluster_5356	17	17	5	12	8	0	0.001647597	0.04937888	0	polysaccharide ABC transporter ATP-binding
Cluster_5864	9	9	0	9	13	3	0.000107686	0.005315681	0	RNA polymerase subunit sigma
Cluster_5952	8	8	0	8	13	4	0.000457666	0.01961147	0	rubrerythrin
Cluster_5784	10	10	0	10	13	2	0.0000202	0.001189034	0	saccharopine dehydrogenase
Cluster_5459	15	15	0	12	13	0	0.000000192	0.0000258	0	selenide, water dikinase SelD
Cluster_5520	14	14	0	11	13	1	0.00000269	0.000213464	0	serine/threonine protein phosphatase
Cluster_6097	7	7	0	7	13	5	0.001647597	0.04937888	0	short-chain dehydrogenase
Cluster_5464	15	15	0	12	13	0	0.000000192	0.0000258	0	short-chain dehydrogenase/reductase
Cluster_6092	7	7	0	7	13	5	0.001647597	0.04937888	0	sugar ABC transporter permease
Cluster_6093	7	7	0	7	13	5	0.001647597	0.04937888	0	sugar ABC transporter permease
Cluster_6096	7	7	0	7	13	5	0.001647597	0.04937888	0	sugar ABC transporter substrate-binding protein
Cluster_5472	15	15	0	12	13	0	0.000000192	0.0000258	0	TetR family transcriptional regulator
Cluster_5676	12	12	0	10	13	2	0.0000202	0.001189034	0	TetR family transcriptional regulator
Cluster_5513	14	13	2	10	11	2	0.001202623	0.0460472	0.044570535	transcriptional regulator
Cluster_5659	12	12	1	9	12	3	0.000982636	0.03850249	0.034180927	transcriptional regulator
Cluster_5521	14	14	0	12	13	0	0.000000192	0.0000258	0	transcriptional regulator
Cluster_5783	10	10	0	10	13	2	0.0000202	0.001189034	0	transcriptional regulator
Cluster_5769	10	10	1	9	12	3	0.000982636	0.03850249	0.034180927	transporter
Cluster_5770	10	10	1	9	12	3	0.000982636	0.03850249	0.034180927	transporter

Cluster_5460	15	15	0	12	13	0	0.000000192	0.0000258	0	tRNA 2-selenouridine(34) synthase MnmH
Cluster_6082	7	7	0	7	13	5	0.001647597	0.04937888	0	two-component sensor histidine kinase
Cluster_6095	7	7	0	7	13	5	0.001647597	0.04937888	0	two-component sensor histidine kinase
Cluster_5358	17	17	5	12	8	0	0.001647597	0.04937888	0	xylose isomerase

[a] P-values from two-sided Fisher's Exact Tests

[b] P-values were corrected using the False Discovery Rate (FDR) approach

Supplemental Table 3.7: List of Gene Ontology (GO) terms enriched in phylogenetic clade A

GOterms	Clade A Presence	Clade B Presence	Clade A Absence	Clade B Absence	P-value [a]	FDR P-value [b]	Odds Ratio [c]	Description	Category
EC:2.2.1.6	26	12	0	12	2.23E-05	1.03E-03	Inf	Acetolactate synthase.	Acetohydroxy acid synthetase. Acetohydroxyacid synthase. Acetolactate pyruvate-lyase (carboxylating). Acetolactic synthetase. Alpha-acetohydroxy acid synthetase. Alpha-acetohydroxyacid synthase. Alpha-acetolactate synthase. Alpha-acetolactate synthetase.
EC:2.4.2.30	13	2	0	10	2.02E-05	1.03E-03	Inf	NAD(+) ADP-ribosyltransferase.	ADP-ribosyltransferase (polymerizing). Poly(adenosine diphosphate ribose) polymerase. Poly(ADP-ribose) synthetase. Poly(ADP-ribose)polymerase.
EC:1.7.1.4	26	12	0	12	2.23E-05	1.03E-03	Inf	Nitrite reductase (NAD(P)H).	Assimilatory nitrite reductase. NAD(P)H:nitrite oxidoreductase. NADH-nitrite oxidoreductase. NADPH-nitrite reductase. Nitrite reductase (reduced nicotinamide adenine dinucleotide (phosphate)).
EC:3.2.1.4	26	14	0	10	1.91E-04	6.86E-03	Inf	Cellulase.	Avicelase. Beta-1,4-endoglucan hydrolase. Beta-1,4-glucanase. Carboxymethyl cellulase. Celludextrinase. Endo-1,4-beta-D-glucanase. Endo-1,4-

									beta-D-glucanohydrolase. Endo-1,4-beta-glucanase. Endoglucanase.
GO:0019645	26	0	0	24	8.23E-15	6.33E-12	Inf	anaerobic electron transport chain	biological_process
GO:0009061	90	58	1	26	1.05E-08	3.10E-06	39.718174882	anaerobic respiration	biological_process
GO:0007166	13	2	0	10	2.02E-05	1.03E-03	Inf	cell surface receptor signaling pathway	biological_process
GO:0022607	26	14	0	10	1.91E-04	6.86E-03	Inf	cellular component assembly	biological_process
GO:1990090	13	2	0	10	2.02E-05	1.03E-03	Inf	cellular response to nerve growth factor stimulus	biological_process
GO:0022416	13	2	0	10	2.02E-05	1.03E-03	Inf	chaeta developmen t	biological_process
GO:0051131	13	0	0	12	1.92E-07	2.46E-05	Inf	chaperone-mediated protein complex assembly	biological_process
GO:0051188	52	35	0	13	2.71E-05	1.23E-03	Inf	cofactor biosynthetic process	biological_process
GO:0006535	78	60	0	12	8.91E-05	3.57E-03	Inf	cysteine biosynthetic process from serine	biological_process

GO:0007010	13	2	0	10	2.02E-05	1.03E-03	Inf	cytoskeleton organization	biological_process
GO:0048473	85	60	6	24	1.14E-04	4.46E-03	5.61223 61498	D-methionine transport	biological_process
GO:0052060	26	12	0	12	2.23E-05	1.03E-03	Inf	evasion or tolerance by symbiont of host-produced nitric oxide	biological_process
GO:0048193	13	2	0	10	2.02E-05	1.03E-03	Inf	Golgi vesicle transport	biological_process
GO:0006783	39	24	0	12	4.79E-05	2.02E-03	Inf	heme biosynthetic process	biological_process
GO:0018130	57	34	21	38	1.48E-03	3.52E-02	3.00994 18314	heterocycle biosynthetic process	biological_process
GO:0042592	26	14	0	10	1.91E-04	6.86E-03	Inf	homeostatic process	biological_process
GO:0046907	13	2	0	10	2.02E-05	1.03E-03	Inf	intracellular transport	biological_process
GO:0009097	247	216	0	12	1.28E-04	4.88E-03	Inf	isoleucine biosynthetic process	biological_process
GO:0032507	13	2	0	10	2.02E-05	1.03E-03	Inf	maintenance of protein location in cell	biological_process
GO:1903047	13	2	0	10	2.02E-05	1.03E-03	Inf	mitotic cell cycle process	biological_process
GO:0032324	91	58	0	26	4.78E-10	2.04E-07	Inf	molybdopterin cofactor biosynthetic	biological_process

GO:0043545	13	0	0	12	1.92E-07	2.46E-05	Inf	process molybdopterin cofactor metabolic process	biological_process
GO:0015718	13	0	0	12	1.92E-07	2.46E-05	Inf	monocarboxylic acid transport	biological_process
GO:2001251	13	2	0	10	2.02E-05	1.03E-03	Inf	negative regulation of chromosome organization	biological_process
GO:0044092	13	2	0	10	2.02E-05	1.03E-03	Inf	negative regulation of molecular function	biological_process
GO:0032074	10	1	3	11	9.83E-04	2.82E-02	29.782730376	negative regulation of nuclease activity	biological_process
GO:0042128	78	36	0	36	7.33E-15	6.33E-12	Inf	nitrate assimilation	biological_process
GO:0043602	13	0	0	12	1.92E-07	2.46E-05	Inf	nitrate catabolic process	biological_process
GO:0006807	448	375	85	117	1.68E-03	3.57E-02	1.6435892193	nitrogen compound metabolic process	biological_process
GO:0006857	52	36	0	12	6.63E-05	2.74E-03	Inf	oligopeptide transport	biological_process
GO:1901362	57	34	21	38	1.48E-03	3.52E-02	3.0099418314	organic cyclic compound biosynthetic process	biological_process

GO:0090407	26	12	0	12	2.23E-05	1.03E-03	Inf	organophosphate biosynthetic process	biological_process
GO:0043436	44	23	8	25	1.16E-04	4.50E-03	5.8625194399	oxoacid metabolic process	biological_process
GO:0045760	13	2	0	10	2.02E-05	1.03E-03	Inf	positive regulation of action potential	biological_process
GO:1904064	13	2	0	10	2.02E-05	1.03E-03	Inf	positive regulation of cation transmembrane transport	biological_process
GO:2000573	13	2	0	10	2.02E-05	1.03E-03	Inf	positive regulation of DNA biosynthetic process	biological_process
GO:0032414	13	2	0	10	2.02E-05	1.03E-03	Inf	positive regulation of ion transmembrane transporter activity	biological_process
GO:0009967	13	2	0	10	2.02E-05	1.03E-03	Inf	positive regulation of signal transduction	biological_process
GO:0010765	13	2	0	10	2.02E-05	1.03E-03	Inf	positive regulation of sodium ion transport	biological_process

GO:0032206	13	2	0	10	2.02E-05	1.03E-03	Inf	positive regulation of telomere maintenance	biological_process
GO:0051347	13	2	0	10	2.02E-05	1.03E-03	Inf	positive regulation of transferase activity	biological_process
GO:0006471	13	2	0	10	2.02E-05	1.03E-03	Inf	protein ADP-ribosylation	biological_process
GO:0008104	13	0	0	12	1.92E-07	2.46E-05	Inf	protein localization	biological_process
GO:2001257	13	2	0	10	2.02E-05	1.03E-03	Inf	regulation of cation channel activity	biological_process
GO:0050793	15	2	11	22	2.75E-04	9.61E-03	14.129161013	regulation of developmental process	biological_process
GO:0010556	24	10	2	14	1.75E-04	6.54E-03	15.736326108	regulation of macromolecule biosynthetic process	biological_process
GO:0051239	13	2	0	10	2.02E-05	1.03E-03	Inf	regulation of multicellular organismal process	biological_process
GO:0043266	13	2	0	10	2.02E-05	1.03E-03	Inf	regulation of potassium ion transport	biological_process
GO:0090364	13	2	0	10	2.02E-05	1.03E-03	Inf	regulation of proteasome assembly	biological_process
GO:1900231	10	1	3	11	9.83E-04	2.82E-02	29.782730376	regulation of single-	biological_process

								species biofilm formation on inanimate substrate	
GO:0032210	13	2	0	10	2.02E-05	1.03E-03	Inf	regulation of telomere maintenance via telomerase	biological_process
GO:0001101	26	3	0	21	3.01E-11	1.44E-08	Inf	response to acid chemical	biological_process
GO:0001666	77	45	1	27	2.62E-09	8.87E-07	45.2805 02553	response to hypoxia	biological_process
GO:0009411	39	27	0	9	7.50E-04	2.46E-02	Inf	response to UV	biological_process
GO:0044765	331	259	163	197	1.31E-03	3.52E-02	1.54381 4943	single-organism transport	biological_process
GO:0050808	39	14	0	22	7.35E-10	2.83E-07	Inf	synapse organization	biological_process
GO:0006313	69	28	243	260	3.57E-05	1.54E-03	2.63255 91458	transposition, DNA-mediated	biological_process
GO:0009099	221	192	0	12	1.26E-04	4.84E-03	Inf	valine biosynthetic process	biological_process
GO:0035220	13	2	0	10	2.02E-05	1.03E-03	Inf	wing disc development	biological_process
GO:0005948	26	12	0	12	2.23E-05	1.03E-03	Inf	acetolactate synthase complex	cellular_component
GO:0005911	13	2	0	10	2.02E-05	1.03E-03	Inf	cell-cell junction	cellular_component

GO:0000781	13	2	0	10	2.02E-05	1.03E-03	Inf	chromosome, telomeric region	cellular_component
GO:0044430	13	2	13	22	1.77E-03	3.71E-02	10.468251167	cytoskeletal part	cellular_component
GO:0016021	9310	8261	4600	4579	8.21E-06	8.53E-04	1.1218308221	integral component of membrane	cellular_component
GO:0031235	39	13	13	35	2.71E-06	2.89E-04	7.8778672084	intrinsic component of the cytoplasmic side of the plasma membrane	cellular_component
GO:0044304	13	2	0	10	2.02E-05	1.03E-03	Inf	main axon	cellular_component
GO:0015630	13	2	0	10	2.02E-05	1.03E-03	Inf	microtubule cytoskeleton	cellular_component
GO:0019008	26	12	0	12	2.23E-05	1.03E-03	Inf	molybdopterin synthase complex	cellular_component
GO:0044799	26	0	0	24	8.23E-15	6.33E-12	Inf	NarGHI complex	cellular_component
GO:0009325	13	0	0	12	1.92E-07	2.46E-05	Inf	nitrate reductase complex	cellular_component
GO:0009344	26	12	0	12	2.23E-05	1.03E-03	Inf	nitrite reductase complex [NAD(P)H]	cellular_component
GO:0005635	13	2	0	10	2.02E-05	1.03E-03	Inf	nuclear envelope	cellular_component
GO:0045202	78	50	13	34	1.41E-04	5.32E-03	4.0461143775	synapse	cellular_component
EC:4.99.1.1	13	0	0	12	1.92E-07	2.46E-05	Inf	Protoporphy	Ferrochelatase. Heme

								rin ferrochelatase.	synthase. Iron chelatase. Protoheme ferro-lyase.
GO:0051538	52	36	0	12	6.63E-05	2.74E-03	Inf	3 iron, 4 sulfur cluster binding	molecular_function
GO:0051539	487	411	59	93	5.85E-04	1.94E-02	1.86660 29273	4 iron, 4 sulfur cluster binding	molecular_function
GO:0003984	26	12	0	12	2.23E-05	1.03E-03	Inf	acetolactate synthase activity	molecular_function
GO:0008810	26	14	0	10	1.91E-04	6.86E-03	Inf	cellulase activity	molecular_function
GO:0009055	272	211	53	89	8.33E-05	3.37E-03	2.16202 88663	electron carrier activity	molecular_function
GO:0004325	13	0	0	12	1.92E-07	2.46E-05	Inf	ferrochelatase activity	molecular_function
GO:0015038	13	0	0	12	1.92E-07	2.46E-05	Inf	glutathione disulfide oxidoreductase activity	molecular_function
GO:0051536	282	214	4	50	3.47E-13	2.23E-10	16.4074 39595	iron-sulfur cluster binding	molecular_function
GO:0043546	13	0	0	12	1.92E-07	2.46E-05	Inf	molybdopterin cofactor binding	molecular_function
GO:0030366	26	12	0	12	2.23E-05	1.03E-03	Inf	molybdopterin synthase activity	molecular_function
GO:0008170	13	1	91	95	1.44E-03	3.52E-02	13.4430 21448	N-methyltransferase	molecular_function

GO:0003950	13	2	0	10	2.02E-05	1.03E-03	Inf	activity NAD+ ADP-ribosyltransferase activity	molecular_function
GO:0008940	26	0	0	24	8.23E-15	6.33E-12	Inf	nitrate reductase activity	molecular_function
GO:0008942	26	12	0	12	2.23E-05	1.03E-03	Inf	nitrite reductase [NAD(P)H] activity	molecular_function
GO:0003676	315	241	231	263	1.59E-03	3.52E-02	1.48751 25213	nucleic acid binding	molecular_function
GO:0061463	10	1	3	11	9.83E-04	2.82E-02	29.7827 30376	O-acetyl-ADP-ribose deacetylase activity	molecular_function
GO:0030674	13	2	0	10	2.02E-05	1.03E-03	Inf	protein binding, bridging	molecular_function
GO:0008428	10	1	3	11	9.83E-04	2.82E-02	29.7827 30376	ribonuclease inhibitor activity	molecular_function
GO:0080007	13	0	0	12	1.92E-07	2.46E-05	Inf	S-nitrosoglutathione reductase activity	molecular_function
GO:0097110	13	2	0	10	2.02E-05	1.03E-03	Inf	scaffold protein binding	molecular_function
GO:0004803	69	28	243	260	3.57E-05	1.54E-03	2.63255 91458	transposase activity	molecular_function
GO:0051082	117	96	0	12	1.07E-04	4.24E-03	Inf	unfolded protein	molecular_function

EC:2.8.1.12	26	12	0	12	2.23E-05	1.03E-03	Inf	binding Molybdopterin synthase.	MPT synthase.
EC:1.7.99.4	26	0	0	24	8.23E-15	6.33E-12	Inf	Nitrate reductase.	Respiratory nitrate reductase.

[a] P-values from two-sided Fisher's Exact Tests

[b] P-values were corrected using the False Discovery Rate (FDR) approach

[c] Odds ratios marked as INF (Infinite) are a result of dividing by zero

Supplemental Table 3.8: List of GO terms enriched in phylogenetic clade B

GOterms	Clade A Presence	Clade B Presence	Clade A Absence	Clade B Absence	P-value [a]	FDR P-value [b]	Odds Ratio	Description	Category
EC:4.2.1.44	5	12	8	0	1.65E-03	3.52E-02	0	Myo-inosose-2 dehydratase.	
EC:4.2.2.12	4	12	9	0	4.58E-04	1.53E-02	0	Xanthan lyase.	
EC:2.7.1.92	5	12	8	0	1.65E-03	3.52E-02	0	5-dehydro-2-deoxygluconokinase.	5-keto-2-deoxygluconokinase.
EC:1.1.1.184	0	17	26	7	3.51E-08	8.44E-06	0	Carbonyl reductase (NADPH).	Aldehyde reductase I. NADPH-dependent carbonyl reductase. Prostaglandin 9-ketoreductase. Xenobiotic ketone reductase.
EC:4.2.99.18	27	36	12	0	1.98E-04	6.97E-03	0	DNA-(apurinic or apyrimidinic site) lyase.	AP endonuclease class I. AP lyase. Deoxyribonuclease (apurinic or apyrimidinic). E.coli endonuclease III. Endodeoxyribonuclease (apurinic or apyrimidinic). Micrococcus luteus UV endonuclease. Phage-T(4) UV endonuclease. Phage-T4 UV endonuclease.
EC:3.3.2.10	1	9	12	3	9.83E-04	2.82E-02	0.034180927188	Soluble epoxide hydrolase.	Arene-oxide hydratase. Aryl epoxide hydrase. Cytosolic epoxide hydrolase. Epoxide hydrase. Epoxide hydratase. Trans-stilbene oxide hydrolase.
GO:0051979	13	22	13	2	1.77E-03	3.71E-02	0.09552693	alginic acid acetylation	biological_process

GO:0042918	99	124	44	8	7.99E-08	1.81E-05	9893 0.146 10762281	alkanesulfonate transport	biological_process
GO:0015837	0	10	13	2	2.02E-05	1.03E-03	0	amine transport	biological_process
GO:0006284	118	120	12	0	4.14E-04	1.42E-02	0	base-excision repair	biological_process
GO:0071360	5	12	8	0	1.65E-03	3.52E-02	0	cellular response to exogenous dsRNA	biological_process
GO:0042632	1	9	12	3	9.83E-04	2.82E-02	0.034 18092 7188	cholesterol homeostasis	biological_process
GO:0033227	5	12	8	0	1.65E-03	3.52E-02	0	dsRNA transport	biological_process
GO:0022617	5	12	8	0	1.65E-03	3.52E-02	0	extracellular matrix disassembly	biological_process
GO:0042462	1	11	12	1	3.27E-05	1.44E-03	0.012 09549 4461	eye photoreceptor cell development	biological_process
GO:0046657	0	15	39	21	2.44E-06	2.84E-04	0	folic acid catabolic process	biological_process
GO:0019310	33	60	32	0	6.00E-12	3.30E-09	0	inositol catabolic process	biological_process
GO:0006691	0	17	26	7	3.51E-08	8.44E-06	0	leukotriene metabolic process	biological_process
GO:0001676	1	9	12	3	9.83E-04	2.82E-02	0.034 18092 7188	long-chain fatty acid metabolic process	biological_process
GO:0061	135	164	203	148	1.61E-03	3.52E-02	0.600	methylglyoxal	biological_process

727							60892993	catabolic process to lactate	
GO:0045488	26	42	26	6	8.22E-05	3.36E-03	0.145 81559127	pectin metabolic process	biological_process
GO:0046839	1	9	12	3	9.83E-04	2.82E-02	0.034 18092 7188	phospholipid dephosphorylation	biological_process
GO:0044854	5	12	8	0	1.65E-03	3.52E-02	0	plasma membrane raft assembly	biological_process
GO:0060355	5	12	8	0	1.65E-03	3.52E-02	0	positive regulation of cell adhesion molecule production	biological_process
GO:1901890	5	12	8	0	1.65E-03	3.52E-02	0	positive regulation of cell junction assembly	biological_process
GO:2000049	5	12	8	0	1.65E-03	3.52E-02	0	positive regulation of cell-cell adhesion mediated by cadherin	biological_process
GO:0045807	5	12	8	0	1.65E-03	3.52E-02	0	positive regulation of endocytosis	biological_process
GO:0034116	5	12	8	0	1.65E-03	3.52E-02	0	positive regulation of heterotypic cell-cell adhesion	biological_process
GO:0032728	5	12	8	0	1.65E-03	3.52E-02	0	positive regulation of	biological_process

GO:1901741	5	12	8	0	1.65E-03	3.52E-02	0	interferon-beta production positive regulation of myoblast fusion	biological_process
GO:0032092	5	12	8	0	1.65E-03	3.52E-02	0	positive regulation of protein binding	biological_process
GO:0048643	5	12	8	0	1.65E-03	3.52E-02	0	positive regulation of skeletal muscle tissue development	biological_process
GO:0032226	5	12	8	0	1.65E-03	3.52E-02	0	positive regulation of synaptic transmission, dopaminergic	biological_process
GO:0034141	5	12	8	0	1.65E-03	3.52E-02	0	positive regulation of toll-like receptor 3 signaling pathway	biological_process
GO:0070528	5	12	8	0	1.65E-03	3.52E-02	0	protein kinase C signaling	biological_process
GO:1903044	5	12	8	0	1.65E-03	3.52E-02	0	protein localization to membrane raft	biological_process
GO:0090181	1	9	12	3	9.83E-04	2.82E-02	0.03418092 7188	regulation of cholesterol metabolic process	biological_process
GO:0002090	5	12	8	0	1.65E-03	3.52E-02	0	regulation of receptor internalization	biological_process

GO:0035023	5	12	8	0	1.65E-03	3.52E-02	0	regulation of Rho protein signal transduction	biological_process
GO:0034143	5	12	8	0	1.65E-03	3.52E-02	0	regulation of toll-like receptor 4 signaling pathway	biological_process
GO:0010044	5	12	8	0	1.65E-03	3.52E-02	0	response to aluminum ion	biological_process
GO:0034976	5	12	8	0	1.65E-03	3.52E-02	0	response to endoplasmic reticulum stress	biological_process
GO:0010842	1	11	12	1	3.27E-05	1.44E-03	0.01209549 4461	retina layer formation	biological_process
GO:0046272	1	9	12	3	9.83E-04	2.82E-02	0.03418092 7188	stilbene catabolic process	biological_process
GO:0003008	1	14	25	10	2.67E-05	1.22E-03	0.03088815 4839	system process	biological_process
GO:0015786	5	12	8	0	1.65E-03	3.52E-02	0	UDP-glucose transport	biological_process
GO:0033559	1	9	12	3	9.83E-04	2.82E-02	0.03418092 7188	unsaturated fatty acid metabolic process	biological_process
GO:0016324	5	12	8	0	1.65E-03	3.52E-02	0	apical plasma membrane	cellular_component
GO:0005913	5	12	8	0	1.65E-03	3.52E-02	0	cell-cell adherens junction	cellular_component
GO:0044291	5	12	8	0	1.65E-03	3.52E-02	0	cell-cell contact zone	cellular_component
GO:0034	5	12	8	0	1.65E-03	3.52E-02	0	centriolar	cellular_component

451								satellite	
GO:0044434	14	23	12	1	8.88E-04	2.82E-02	0.053616061358	chloroplast part	cellular_component
GO:0030864	5	12	8	0	1.65E-03	3.52E-02	0	cortical actin cytoskeleton	cellular_component
GO:0016600	5	12	8	0	1.65E-03	3.52E-02	0	flotillin complex	cellular_component
GO:0005925	5	12	8	0	1.65E-03	3.52E-02	0	focal adhesion	cellular_component
GO:0019814	1	9	12	3	9.83E-04	2.82E-02	0.034180927188	immunoglobulin complex	cellular_component
GO:0030027	5	12	8	0	1.65E-03	3.52E-02	0	lamellipodium	cellular_component
GO:0009526	40	47	12	1	2.02E-03	4.22E-02	0.072385409793	plastid envelope	cellular_component
GO:0001931	5	12	8	0	1.65E-03	3.52E-02	0	uropod	cellular_component
EC:3.1.21.2	1	12	12	0	2.69E-06	2.89E-04	0	Deoxyribonuclease IV.	Deoxyribonuclease IV (phage T4-induced). Endodeoxyribonuclease IV (phage T(4)-induced). Endodeoxyribonuclease IV (phage T4-induced). Endonuclease II. Endonuclease IV.
GO:0047590	5	12	8	0	1.65E-03	3.52E-02	0	5-dehydro-2-deoxygluconokinase activity	molecular_function
GO:0042959	42	48	10	0	1.29E-03	3.52E-02	0	alkanesulfonate transporter activity	molecular_function
GO:0016837	1	9	12	3	9.83E-04	2.82E-02	0.034180927188	carbon-oxygen lyase activity, acting on	molecular_function

								polysaccharides	
GO:0004090	0	17	26	7	3.51E-08	8.44E-06	0	carbonyl reductase (NADPH) activity	molecular_function
GO:0008833	1	12	12	0	2.69E-06	2.89E-04	0	deoxyribonuclease IV (phage-T4-induced) activity	molecular_function
GO:0003906	27	36	12	0	1.98E-04	6.97E-03	0	DNA-(apurinic or apyrimidinic site) lyase activity	molecular_function
GO:0004301	1	9	12	3	9.83E-04	2.82E-02	0.034180927188	epoxide hydrolase activity	molecular_function
GO:0016817	39	67	39	5	2.77E-09	8.87E-07	0.075966522498	hydrolase activity, acting on acid anhydrides	molecular_function
GO:0042577	1	9	12	3	9.83E-04	2.82E-02	0.034180927188	lipid phosphatase activity	molecular_function
GO:0005537	4	12	9	0	4.58E-04	1.53E-02	0	mannose binding	molecular_function
GO:0050114	5	12	8	0	1.65E-03	3.52E-02	0	myo-inosose-2 dehydratase activity	molecular_function
GO:0016705	48	68	30	4	1.15E-06	1.42E-04	0.095515329635	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	molecular_function

GO:0071713	0	15	39	21	2.44E-06	2.84E-04	0	para-aminobenzoyl-glutamate hydrolase activity	molecular_function
GO:0008081	79	84	12	0	3.66E-04	1.27E-02	0	phosphoric diester hydrolase activity	molecular_function
GO:0002020	5	12	8	0	1.65E-03	3.52E-02	0	protease binding	molecular_function
GO:0046982	44	61	47	23	1.21E-03	3.44E-02	0.35516123212	protein heterodimerization activity	molecular_function
GO:0102483	0	12	39	24	4.79E-05	2.02E-03	0	scopolin beta-glucosidase activity	molecular_function
GO:0031132	0	12	13	0	1.92E-07	2.46E-05	0	serine 3-dehydrogenase activity	molecular_function
GO:0015643	1	9	12	3	9.83E-04	2.82E-02	0.0341809271 88	toxic substance binding	molecular_function
GO:0005460	5	12	8	0	1.65E-03	3.52E-02	0	UDP-glucose transmembrane transporter activity	molecular_function
GO:0047492	4	12	9	0	4.58E-04	1.53E-02	0	xanthan lyase activity	molecular_function
EC:1.1.1.276	0	12	13	0	1.92E-07	2.46E-05	0	Serine 3-dehydrogenase (NADP(+)).	Serine 3-dehydrogenase.

[a] P-values from two-sided Fisher's Exact Tests

[b] P-values were corrected using the False Discovery Rate (FDR) approach

APPENDIX C

Appendix C-1: Growth data (CFU/ml) for *B. cereus* group isolates grown in BHI and SMB at 6°C on days 0, 14, and 21.

	Day 0		Day 14		Day 21	
Isolate ID (Replicate)	BHI	SMB	BHI	SMB	BHI	SMB
FSL M7-0669 (1)	2.45E+02	2.86E+02	1.51E+03	7.17E+04	1.37E+05	3.21E+06
FSL M7-0669 (2)	3.68E+02	5.31E+02	3.55E+04	6.50E+04	2.15E+05	2.32E+05
FSL M7-0669 (3)	4.70E+02	4.70E+02	1.54E+04	2.04E+01	1.30E+05	4.09E+01
FSL W8-0169 (1)	1.00E+03	1.84E+02	6.13E+01	2.25E+02	0.00E+00[a]	0.00E+00[a]
FSL W8-0169 (2)	1.17E+03	2.04E+02	1.02E+02	3.07E+02	0.00E+00[a]	0.00E+00[a]
FSL W8-0169 (3)	1.29E+03	1.84E+02	0.00E+00[a]	6.13E+01	4.09E+01	2.04E+01
FSL H7-0683 (1)	6.54E+02	3.27E+02	6.13E+01	0.00E+00[a]	1.66E+03	2.04E+01
FSL H7-0683 (2)	5.93E+02	3.68E+02	2.04E+02	0.00E+00[a]	8.38E+02	2.04E+01
FSL H7-0683 (3)	5.93E+02	5.11E+02	0.00E+00[a]	2.04E+01	1.02E+02	0.00E+00[a]
FSL H7-0926 (1)	2.04E+02	2.25E+02	0.00E+00[a]	4.09E+01	2.04E+01	6.13E+01
FSL H7-0926 (2)	3.88E+02	3.27E+02	0.00E+00[a]	6.13E+01	0.00E+00[a]	2.04E+01
FSL H7-0926 (3)	2.86E+02	3.07E+02	0.00E+00[a]	4.09E+01	0.00E+00[a]	8.38E+02
FSL M7-1219 (1)	4.09E+01	0.00E+00[a]	1.84E+02	2.39E+03	1.39E+03	1.05E+04
FSL M7-1219 (2)	4.09E+01	0.00E+00[a]	9.61E+02	0.00E+00[a]	2.68E+03	0.00E+00[a]
FSL M7-1219 (3)	0.00E+00[a]	0.00E+00[a]	8.18E+02	0.00E+00[a]	1.39E+03	0.00E+00[a]
FSL H8-0485 (1)	2.29E+03	1.17E+03	4.84E+04	0.00E+00[a]	1.50E+05	0.00E+00[a]
FSL H8-0485 (2)	2.43E+03	1.23E+03	3.79E+04	0.00E+00[a]	1.56E+05	0.00E+00[a]
FSL H8-0485 (3)	2.55E+03	1.21E+03	2.65E+04	2.04E+01	1.30E+05	0.00E+00[a]
FSL H8-0492 (1)	1.14E+03	4.91E+02	1.05E+05	2.04E+01	7.06E+04	1.43E+02
FSL H8-0492 (2)	1.00E+03	4.70E+02	5.30E+04	4.09E+01	1.37E+05	2.04E+01
FSL H8-0492 (3)	1.00E+03	4.70E+02	4.03E+03	2.04E+01	1.71E+05	0.00E+00[a]
FSL R5-0708 (1)	1.59E+03	1.14E+03	1.99E+04	1.31E+03	1.63E+05	6.20E+03
FSL R5-0708 (2)	1.37E+03	1.29E+03	1.08E+03	3.30E+04	7.33E+04	2.38E+05
FSL R5-0708 (3)	1.64E+03	1.12E+03	7.52E+03	8.18E+01	1.22E+05	2.04E+01
FSL M7-0109 (1)	8.79E+02	7.15E+02	4.51E+04	7.15E+02	2.16E+05	3.37E+03
FSL M7-0109 (2)	6.95E+02	2.66E+02	1.30E+05	5.52E+02	2.00E+05	2.33E+03
FSL M7-0109 (3)	6.74E+02	4.50E+02	1.21E+05	2.25E+02	2.04E+05	8.18E+02
FSL W7-1108 (1)	3.47E+02	1.64E+02	8.71E+04	0.00E+00[a]	4.98E+04	0.00E+00[a]
FSL W7-1108 (2)	4.50E+02	3.07E+02	5.74E+04	6.13E+01	1.54E+04	0.00E+00[a]
FSL W7-1108 (3)	4.70E+02	3.27E+02	6.72E+04	0.00E+00[a]	1.61E+03	1.23E+02
FSL M8-0091 (1)	8.38E+02	7.77E+02	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]
FSL M8-0091 (2)	1.12E+03	5.72E+02	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]
FSL M8-0091 (3)	8.38E+02	4.70E+02	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]
FSL J3-0113 (1)	4.05E+03	5.11E+02	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]
FSL J3-0113 (2)	3.94E+03	7.97E+02	0.00E+00[a]	2.04E+01	0.00E+00[a]	0.00E+00[a]
FSL J3-0113 (3)	1.86E+03	7.36E+02	6.13E+01	2.04E+01	0.00E+00[a]	0.00E+00[a]
FSL J3-0123 (1)	4.70E+02	2.66E+02	3.68E+02	8.18E+01	3.92E+03	4.09E+01
FSL J3-0123 (2)	4.70E+02	3.27E+02	3.35E+03	6.13E+01	1.19E+05	2.04E+01
FSL J3-0123 (3)	6.13E+02	1.84E+02	1.85E+04	2.04E+01	1.09E+05	2.04E+01
FSL E2-0214 (1)	1.02E+03	7.56E+02	8.39E+04	4.09E+01	3.10E+05	0.00E+00[a]
FSL E2-0214 (2)	7.97E+02	6.95E+02	5.23E+04	0.00E+00[a]	1.71E+05	0.00E+00[a]
FSL E2-0214 (3)	7.77E+02	4.29E+02	4.04E+04	0.00E+00[a]	1.88E+05	0.00E+00[a]
FSL K6-1030 (1)	1.84E+02	4.29E+02	0.00E+00[a]	2.04E+01	0.00E+00[a]	0.00E+00[a]
FSL K6-1030 (2)	2.86E+02	2.66E+02	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]
FSL K6-1030 (3)	2.86E+02	2.86E+02	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]
FSL W8-0483 (1)	8.99E+02	5.93E+02	4.09E+01	1.02E+02	0.00E+00[a]	0.00E+00[a]
FSL W8-0483 (2)	2.76E+03	4.91E+02	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]	0.00E+00[a]

FSL W8-0483 (3)	5.93E+02	4.91E+02	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]
FSL M8-0117 (1)	6.13E+01	1.43E+02	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]
FSL M8-0117 (2)	8.18E+01	2.45E+02	0.00E+00 [a]	0.00E+00 [a]	2.04E+01	0.00E+00 [a]
FSL M8-0117 (3)	5.52E+02	1.23E+02	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	4.09E+01
FSL W8-0050 (1)	7.36E+02	7.15E+02	0.00E+00 [a]	2.04E+01	0.00E+00 [a]	0.00E+00 [a]
FSL W8-0050 (2)	8.38E+02	6.34E+02	0.00E+00 [a]	4.09E+01	0.00E+00 [a]	0.00E+00 [a]
FSL W8-0050 (3)	9.61E+02	6.13E+02	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	4.09E+01
FSL K6-0069 (1)	1.66E+03	5.31E+02	0.00E+00 [a]	0.00E+00 [a]	2.04E+01	4.09E+01
FSL K6-0069 (2)	1.45E+03	4.70E+02	8.18E+01	0.00E+00 [a]	0.00E+00 [a]	4.09E+01
FSL K6-0069 (3)	1.10E+03	4.50E+02	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	4.09E+01
FSL R5-0811 (1)	3.19E+03	1.00E+03	0.00E+00 [a]	6.13E+01	0.00E+00 [a]	2.04E+01
FSL R5-0811 (2)	2.31E+03	1.08E+03	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	2.04E+01
FSL R5-0811 (3)	2.21E+03	1.12E+03	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]
FSL M8-0473 (1)	9.81E+02	9.61E+02	2.04E+01	0.00E+00 [a]	0.00E+00 [a]	2.04E+01
FSL M8-0473 (2)	9.20E+02	6.95E+02	0.00E+00 [a]	0.00E+00 [a]	2.04E+01	2.04E+01
FSL M8-0473 (3)	9.20E+02	5.52E+02	2.04E+01	0.00E+00 [a]	0.00E+00 [a]	2.04E+01
FSL W8-0268 (1)	2.62E+03	1.08E+03	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	2.04E+01
FSL W8-0268 (2)	3.43E+03	1.19E+03	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]
FSL W8-0268 (3)	2.66E+03	9.61E+02	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]
FSL H8-0534 (1)	4.09E+01	2.25E+02	2.04E+01	0.00E+00 [a]	0.00E+00 [a]	2.04E+01
FSL H8-0534 (2)	1.02E+02	2.86E+02	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]
FSL H8-0534 (3)	6.13E+01	2.66E+02	0.00E+00 [a]	0.00E+00 [a]	0.00E+00 [a]	2.04E+01

[a]These values (0.00E+0) are below the limit of detection. The detection limit for this study was 10 CFU/ml. As such, the true values of these counts are < 10 CFU/ml

Appendix C-2: Growth data for *B. cereus* group isolates grown in BHI and SMB at 6°C where day 0 is given in CFU/ml and days 14 and 21 are expressed as log CFU/ml change from day 0.

	BHI Rep 1			BHI Rep 2			BHI Rep 3			SMB Rep 1			SMB Rep 2			SMB Rep 3		
FSL Isolate	Day 0	Day 14	Day 21	Day 0	Day 14	Day 21	Day 0	Day 14	Day 21	Day 0	14	21	Day 0	14	21	Day 0	14	21
M7-0669	2.45E+2	3.1	5.14	3.68E+2	4.55	5.33	4.70E+2	4.17	5.11	2.86E+2	4.85	6.51	5.31E+2	4.81	5.36	4.70E+2	-2.65	-2.63
W8-0169	1.00E+3	-2.97	≤ -2*	1.17E+3	-3.03	≤ -2.07*	1.29E+3	≤ -2.11*	-3.1	1.84E+2	1.61	1.26*	2.04E+2	2.01	≤ -1.31*	1.84E+2	-2.09	-2.21
H7-0683	6.54E+2	-2.77	3	5.93E+2	-2.59	2.39	5.93E+2	≤ -1.77*	-2.69	3.27E+2	≤ -1.51*	-2.49	3.68E+2	≤ -1.57*	-2.54	5.11E+2	-2.69	-2.71
H7-0926	2.04E+2	≤ -1.31*	-2.26	3.88E+2	≤ -1.59*	≤ -1.59*	2.86E+2	≤ -1.46*	≤ -1.46*	2.25E+2	-2.27	-2.21	3.27E+2	-2.42	-2.49	3.07E+2	-2.43	2.73
M7-1219	4.09E+1	2.16	3.13	4.09E+1	2.96	3.42	0.00E+0	2.91	3.14	0.00E+0[a]	3.39	4.02	0.00E+0[a]	0[a]	0[a]	0.00E+0[a]	0[a]	0[a]
H8-0485	2.29E+3	4.66	5.17	2.43E+3	4.55	5.19	2.55E+3	4.38	5.11	1.17E+3	≤ -2.07*	≤ -2.07*	1.23E+3	≤ -2.09*	≤ -2.09*	1.21E+3	-3.08	≤ -2.08*
H8-0492	1.14E+3	5.02	4.84	1.00E+3	4.72	5.13	1.00E+3	3.48	5.23	4.91E+2	-2.67	-2.54	4.70E+2	-2.63	-2.65	4.70E+2	-2.65	≤ -1.67*
R5-0708	1.59E+3	4.26	5.21	1.37E+3	-2.46	4.86	1.64E+3	3.77	5.08	1.14E+3	2.23	3.7	1.29E+3	4.5	5.4	1.12E+3	-3.02	-3.04
M7-0109	8.79E+2	4.65	5.33	6.95E+2	5.11	5.3	6.74E+2	5.08	5.31	7.15E+2	≤ -1.85*	3.42	2.66E+2	2.46	3.31	4.50E+2	-2.35	2.57
W7-1108	3.47E+2	4.94	4.69	4.50E+2	4.76	4.17	4.70E+2	4.82	3.06	1.64E+2	≤ -1.21*	≤ -1.21*	3.07E+2	-2.39	≤ -1.49*	3.27E+2	≤ -1.51*	-2.31
M8-0091	8.38E+2	≤ -1.92*	≤ -1.92*	1.12E+3	≤ -2.05*	≤ -2.05*	8.38E+2	≤ -1.92*	≤ -1.92*	7.77E+2	≤ -1.89*	≤ -1.89*	5.72E+2	≤ -1.76*	≤ -1.76*	4.70E+2	≤ -1.67*	≤ -1.67*
J3-0113	4.05E+3	≤ -2.61*	≤ -2.61*	3.94E+3	≤ -2.60*	≤ -2.60*	1.86E+3	≤ -3.25	≤ -2.27*	5.11E+2	≤ -1.71*	≤ -1.71*	7.97E+2	-2.89	≤ -1.90*	7.36E+2	-2.85	≤ -1.87*
J3-0123	4.70E+2	-2.01	3.54	4.70E+2	3.46	5.07	6.13E+2	4.25	5.03	2.66E+2	-2.27	-2.35	3.27E+2	-2.42	-2.49	1.84E+2	-2.21	-2.21
E2-0214	1.02E+3	4.92	5.49	7.97E+2	4.71	5.23	7.77E+2	4.6	5.35	7.56E+2	-2.85	≤ -1.88*	6.95E+2	≤ -1.84*	≤ -1.84*	4.29E+2	≤ -1.63*	≤ -1.63*
K6-1030	1.84E+2	≤ -1.26*	≤ -1.26*	2.86E+2	≤ -1.46*	≤ -1.46*	2.86E+2	≤ -1.46*	≤ -1.46*	4.29E+2	-2.61	≤ -1.63*	2.66E+2	≤ -1.42*	≤ -1.42*	2.86E+2	≤ -1.46*	≤ -1.46*
W8-0483	8.99E+2	-2.93	≤ -1.95*	2.76E+2	≤ -2.44*	≤ -2.44*	5.93E+2	≤ -1.77*	≤ -1.77*	5.93E+2	-2.69	≤ -1.77*	4.91E+2	≤ -1.69*	≤ -1.69*	4.91E+2	≤ -1.69*	≤ -1.69*
M8-0117	6.13E+1	≤ -0.78*	≤ -0.78*	8.18E+1	≤ -0.91*	-1.79	5.52E+2	≤ -1.74*	≤ -1.74*	1.43E+2	≤ -1.16*	≤ -1.16*	2.45E+2	≤ -1.39*	≤ -1.39*	1.23E+2	≤ -1.09*	-1.91
W8-0050	7.36E+2	≤ -1.87*	≤ -1.87*	8.38E+2	≤ -1.92*	≤ -1.92*	9.61E+2	≤ -1.98*	≤ -1.98*	7.15E+2	-2.84	≤ -1.85*	6.34E+2	-2.77	≤ -1.80*	6.13E+2	≤ -1.79*	-2.76
K6-0069	1.66E+3	≤ -2.22*	-3.21	1.45E+3	-3.14	≤ -2.16*	1.10E+3	≤ -2.04*	≤ -2.04*	5.13E+2	≤ -1.71*	-2.69	4.70E+2	≤ -1.67*	-2.63	4.50E+2	≤ -1.65*	-2.61
R5-0811	3.19E+3	≤ -2.50*	≤ -2.50*	2.31E+3	≤ -2.36*	≤ -2.36*	2.21E+3	≤ -2.34*	≤ -2.34*	1.00E+3	-2.97	-2.99	1.08E+3	≤ -2.03*	-3.03	1.12E+3	≤ -2.05*	≤ -2.05*
M8-0473	9.81E+2	-2.98	≤ -1.99*	9.20E+2	≤ -1.96*	-2.95	9.20E+2	-2.95	≤ -1.96*	9.61E+2	≤ -1.98*	-2.97	6.95E+2	≤ -1.84*	-2.83	5.52E+2	≤ -1.74*	-2.73
W8-0268	2.62E+3	≤ -2.42*	≤ -2.42*	3.43E+3	≤ -2.54*	≤ -2.54*	2.66E+3	≤ -2.42*	≤ -2.42*	1.08E+3	≤ -2.03*	-3.03	1.19E+3	≤ -2.08*	≤ -2.08*	9.61E+2	≤ -1.98*	≤ -1.98*
H8-0534	4.09E+1	-1.31	≤ -0.61*	1.02E+2	≤ -1.01*	≤ -1.01*	6.13E+1	≤ -0.79*	≤ -0.78*	2.25E+2	≤ -1.35*	-2.31	2.86E+2	≤ -1.46*	≤ -1.46*	2.26E+2	≤ -1.35*	-2.39

*These values were below the limit of detection (count of 0, meaning < 10 CFU/ml). Because we cannot assign a true reduction to these counts, they are marked as "at least" (≤) this given reduction.

[a]These values (0.00E+0 and 0) are below the limit of detection. The limit of detection in this study was 10 CFU/ml, as such, these values are < 10 CFU/ml.

Appendix C-3: *B. cereus* group isolates were screened for growth on BHI agar plates. Plates were incubated at 6°C for 21 days.

Isolate	Screen Result
FSL M7-0669	Grows
FSL W8-0169	Grows
FSL H7-0683	Grows
FSL H7-0926	Grows
FSL M7-1219	Grows
FSL H8-0485	Grows
FSL H8-0492	Grows
FSL R5-0708	Grows
FSL M7-0109	Grows
FSL W7-1108	Grows
FSL M8-0091	Variable[a]
FSL J3-0113	Variable[a]
FSL J3-0123	Grows
FSL E2-0214	Was not screened[b]
FSL K6-1030	Was not screened[b]
FSL W8-0483	Doesn't grow
FSL M8-0117	Doesn't grow
FSL W8-0050	Doesn't grow
FSL K6-0069	Doesn't grow
FSL R5-0811	Doesn't grow
FSL M8-0473	Doesn't grow
FSL W8-0268	Was not screened[c]
FSL H8-0534	Doesn't grow

[a] Growth was tested on 3 different plates. These isolates showed growth twice and did not grow in the third replicate. As such, their growth is marked as variable.
[b] FSL K6-1030 and FSL E2-0214 were identified as new *rpoB* allelic types following this screening experiment and were therefore included in the in depth testing.
[c] FSL W8-0268 was not included in the screen. Another isolate (FSL W8-0824) with the same *rpoB* allelic type had been screened and did not grow.

Supplemental Table 4.1: Twelve protein domains previous associated with growth at low temperatures used for Hidden Markov Model (HMM) analyses.

Query	Accession	Description	Length
Caps_synth_CapC	PF1402.5	Capsule biosynthesis CapC	119 aa
CSD	PF00313.21	Cold-shock' DNA-binding domain	66 aa
DEADboxA	PF12343.7	Cold shock protein DEAD box A	69 aa
DEAD	PF00270.28	DEAD/DEAH box helicase	176 aa
DnaJ	PF00226.30	DnaJ domain	63 aa
FA_desaturase_2	PF03405.13	Fatty acid desaturase	326 aa
FA_desaturase	PF00487.23	Fatty acid desaturase	254 aa
FA_hydroxylase	PF04116.12	Fatty acid hydroxylase superfamily	133 aa
LtrA	PF06772.10	Bacterial low temperature requirement A protein (LtrA)	353 aa
Peptidase_S11	PF00768.19	D-alanyl-D-alanine carboxypeptidase	241 aa
RecA	PF00154.20	recA bacterial DNA recombination protein	263 aa
YdjO	PF14169.5	Cold-inducible protein YdjO	59 aa

Supplemental Table 4.2: OrthoMCL results: List of genes encoded in the genomes of *B. cereus* group isolates that can grow at 6°C

Cluster	Number of Genes	Number of Taxa	Presence Cold Growers	Absence Cold Growers	Presence Non-Cold Growers	Absence Non-Cold Growers	P-value [a]	FDR P-value [b]	Odds Ratio [c]	Products
Cluster_5932	6	6	6	3	0	14	0.00083212	0.02670618	Inf	2'-5' RNA ligase
Cluster_5930	6	6	6	3	0	14	0.00083212	0.02670618	Inf	hypothetical protein
Cluster_5931	6	6	6	3	0	14	0.00083212	0.02670618	Inf	hypothetical protein
Cluster_5933	6	6	6	3	0	14	0.00083212	0.02670618	Inf	saccharopine dehydrogenase
Cluster_5934	6	6	6	3	0	14	0.00083212	0.02670618	Inf	transcriptional regulator
Cluster_5661	7	7	7	2	0	14	0.00046845	0.01129584	Inf	carboxymuconolactone decarboxylase family
Cluster_5663	7	7	7	2	0	14	0.00046845	0.01129584	Inf	hypothetical protein
Cluster_5279	8	8	7	2	1	13	0.00146268	0.03162802	34.3948411	cold-shock protein
Cluster_5269	8	8	7	2	1	13	0.00146268	0.03162802	34.3948411	hypothetical protein
Cluster_5277	8	8	7	2	1	13	0.00146268	0.03162802	34.3948411	hypothetical protein
Cluster_5301	8	8	7	2	1	13	0.00146268	0.03162802	34.3948411	hypothetical protein
Cluster_5374	8	8	7	2	1	13	0.00146268	0.03162802	34.3948411	hypothetical protein
Cluster_5375	8	8	7	2	1	13	0.0010	0.031628	34.39	hypothetical protein

							46268	02	48411
									8
									34.39
							0.0010	0.031628	48411
Cluster_5376	8	8	7	2	1	13	46268	02	8
									34.39
							0.0010	0.031628	48411
Cluster_5377	8	8	7	2	1	13	46268	02	8
									34.39
							0.0010	0.031628	48411
Cluster_5378	8	8	7	2	1	13	46268	02	8
									34.39
							0.0010	0.031628	48411
Cluster_5380	8	8	7	2	1	13	46268	02	8
									34.39
							0.0010	0.031628	48411
Cluster_5381	8	8	7	2	1	13	46268	02	8
									34.39
							0.0010	0.031628	48411
Cluster_5384	8	8	7	2	1	13	46268	02	8
									34.39
							0.0010	0.031628	48411
Cluster_5388	8	8	7	2	1	13	46268	02	8
									34.39
							0.0010	0.031628	48411
Cluster_5391	8	8	7	2	1	13	46268	02	8
									34.39
							0.0010	0.031628	48411
Cluster_5389	8	8	7	2	1	13	46268	02	8
									34.39
							0.0010	0.031628	48411
Cluster_5390	8	8	7	2	1	13	46268	02	8
									34.39
							0.0010	0.031628	48411
Cluster_5382	8	8	7	2	1	13	46268	02	8
Cluster_5379	8	8	7	2	1	13	0.0010	0.031628	34.39

hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
hypothetical protein
NUDIX hydrolase
polysaccharide deacetylase
RNA polymerase subunit sigma-24
transcriptional regulator

							46268	02	484118	
Cluster_4940	10	9	8	1	1	13	0.000155411	0.01129584	65.557162	alpha/beta hydrolase
Cluster_5086	9	9	8	1	1	13	0.000155411	0.01129584	65.557162	cyclic pyranopterin monophosphate synthase
Cluster_5189	9	9	8	1	1	13	0.000155411	0.01129584	65.557162	hypothetical protein
Cluster_5190	9	9	8	1	1	13	0.000155411	0.01129584	65.557162	hypothetical protein
Cluster_5194	9	9	8	1	1	13	0.000155411	0.01129584	65.557162	hypothetical protein
Cluster_5195	9	9	8	1	1	13	0.000155411	0.01129584	65.557162	hypothetical protein
Cluster_5193	9	9	8	1	1	13	0.000155411	0.01129584	65.557162	phosphatase
Cluster_5187	9	9	8	1	1	13	0.000155411	0.01129584	65.557162	recombinase RecQ
Cluster_4939	10	10	8	1	2	12	0.000728105	0.02375352	36.210825355	acetyltransferase
Cluster_4945	10	10	8	1	2	12	0.000728105	0.02375352	36.210825355	acetyltransferase
Cluster_4972	10	10	8	1	2	12	0.000728105	0.02375352	36.210825355	acetyltransferase
Cluster_4979	10	10	8	1	2	12	0.0007	0.023753	36.21	cold-shock protein

							28105	52	082535	
									36.21	
Cluster_4978	10	10	8	1	2	12	0.0007 28105	0.023753 52	082535	copper oxidase
									36.21	
Cluster_4967	10	10	8	1	2	12	0.0007 28105	0.023753 52	082535	cupin
									36.21	
Cluster_4953	10	10	8	1	2	12	0.0007 28105	0.023753 52	082535	damage-inducible protein DinB
									36.21	
Cluster_4968	10	10	8	1	2	12	0.0007 28105	0.023753 52	082535	dihydrolipoamide dehydrogenase
									36.21	
Cluster_4925	10	10	8	1	2	12	0.0007 28105	0.023753 52	082535	DNA mismatch repair protein MutT
									36.21	
Cluster_4922	10	10	8	1	2	12	0.0007 28105	0.023753 52	082535	flavoprotein
									36.21	
Cluster_4937	10	10	8	1	2	12	0.0007 28105	0.023753 52	082535	glyoxalase
									36.21	
Cluster_4966	10	10	8	1	2	12	0.0007 28105	0.023753 52	082535	GNAT family acetyltransferases
									36.21	
Cluster_4934	10	10	8	1	2	12	0.0007 28105	0.023753 52	082535	group-specific protein
									36.21	
Cluster_4955	10	10	8	1	2	12	0.0007 28105	0.023753 52	082535	group-specific protein
									36.21	
Cluster_4938	10	10	8	1	2	12	0.0007 28105	0.023753 52	082535	hypothetical protein
Cluster_4942	10	10	8	1	2	12	0.0007	0.023753	36.21	hypothetical protein

							28105	52	082535	
									36.21	
							0.0007	0.023753	082535	
Cluster_4950	10	10	8	1	2	12	28105	52	36.21	hypothetical protein
							0.0007	0.023753	082535	
Cluster_4951	10	10	8	1	2	12	28105	52	36.21	hypothetical protein
							0.0007	0.023753	082535	
Cluster_4962	10	10	8	1	2	12	28105	52	36.21	hypothetical protein
							0.0007	0.023753	082535	
Cluster_4970	10	10	8	1	2	12	28105	52	36.21	hypothetical protein
							0.0007	0.023753	082535	
Cluster_4971	10	10	8	1	2	12	28105	52	36.21	hypothetical protein
							0.0007	0.023753	082535	
Cluster_4974	10	10	8	1	2	12	28105	52	36.21	hypothetical protein
							0.0007	0.023753	082535	
Cluster_4975	10	10	8	1	2	12	28105	52	36.21	hypothetical protein
							0.0007	0.023753	082535	
Cluster_4976	10	10	8	1	2	12	28105	52	36.21	hypothetical protein
							0.0007	0.023753	082535	
Cluster_4949	10	10	8	1	2	12	28105	52	36.21	methyltransferase
							0.0007	0.023753	082535	molybdopterin synthase
Cluster_4944	10	10	8	1	2	12	28105	52	36.21	sulfur carrier subunit
							0.0007	0.023753	082535	N-acetylmuramoyl-L-
Cluster_4936	10	10	8	1	2	12	28105	52	36.21	alanine amidase
Cluster_4918	10	10	8	1	2	12	0.0007	0.023753	36.21	PbsX family

							28105	52	082535	transcriptional regulator
Cluster_4946	10	10	8	1	2	12	0.00028105	0.02375352	36.2108253	S-layer protein
Cluster_4969	10	10	8	1	2	12	0.00028105	0.02375352	36.2108253	TetR family transcriptional regulator
Cluster_4965	10	10	8	1	2	12	0.00028105	0.02375352	36.2108253	thiamine pyrophosphate-binding protein
Cluster_4866	10	10	8	1	2	12	0.00028105	0.02375352	36.2108253	transcriptional regulator
Cluster_4919	10	10	8	1	2	12	0.00028105	0.02375352	36.2108253	Virginiamycin B lyase
Cluster_4754	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	6-phosphogluconolactonase
Cluster_4736	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	ABC transporter permease
Cluster_4779	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	ABC transporter permease
Cluster_4701	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	acetyltransferase
Cluster_4717	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	acetyltransferase
Cluster_4730	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	acetyltransferase
Cluster_4768	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	acetyltransferase
Cluster_4790	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	acetyltransferase
Cluster_4711	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	alkyl hydroperoxide reductase
Cluster_4703	11	11	9	0	2	12	6.73E-	0.005280	Inf	alpha/beta hydrolase

							05	144		
							6.73E-	0.005280		aminoglycoside
Cluster_4745	11	11	9	0	2	12	05	144	Inf	phosphotransferase
							6.73E-	0.005280		aminoglycoside
Cluster_4794	11	11	9	0	2	12	05	144	Inf	phosphotransferase
							6.73E-	0.005280		antibiotic biosynthesis
Cluster_4741	11	11	9	0	2	12	05	144	Inf	monooxygenase
							6.73E-	0.005280		AraC family
Cluster_4739	11	11	9	0	2	12	05	144	Inf	transcriptional regulator
							6.73E-	0.005280		AraC family
Cluster_4793	11	11	9	0	2	12	05	144	Inf	transcriptional regulator
							6.73E-	0.005280		capsular biosynthesis
Cluster_4803	11	11	9	0	2	12	05	144	Inf	protein
							6.73E-	0.005280		
Cluster_4749	11	11	9	0	2	12	05	144	Inf	cell surface protein
							6.73E-	0.005280		
Cluster_4764	11	11	9	0	2	12	05	144	Inf	cell wall anchor protein
							6.73E-	0.005280		chromosome segregation
Cluster_4778	11	11	9	0	2	12	05	144	Inf	protein
							6.73E-	0.005280		
Cluster_4759	11	11	9	0	2	12	05	144	Inf	competence protein ComF
							6.73E-	0.005280		
Cluster_4731	11	11	9	0	2	12	05	144	Inf	competence protein ComG
							6.73E-	0.005280		
Cluster_4732	11	11	9	0	2	12	05	144	Inf	competence protein ComG
							6.73E-	0.005280		
Cluster_4733	11	11	9	0	2	12	05	144	Inf	competence protein ComG
							6.73E-	0.005280		
Cluster_4734	11	11	9	0	2	12	05	144	Inf	competence protein ComG
							6.73E-	0.005280		
Cluster_4719	11	11	9	0	2	12	05	144	Inf	competence protein
							6.73E-	0.005280		cytochrome C oxidase
Cluster_4756	11	11	9	0	2	12	05	144	Inf	subunit II
							6.73E-	0.005280		DNA mismatch repair
Cluster_4781	11	11	9	0	2	12	05	144	Inf	protein MutT
Cluster_4735	11	11	9	0	2	12	6.73E-	0.005280	Inf	DNA recombination

							05	144		protein RecO
							6.73E-	0.005280		glutamine
Cluster_4769	11	11	9	0	2	12	05	144	Inf	amidotransferase
							6.73E-	0.005280		
Cluster_4726	11	11	9	0	2	12	05	144	Inf	glycosyl transferase
							6.73E-	0.005280		
Cluster_4727	11	11	9	0	2	12	05	144	Inf	glycosyl transferase
							6.73E-	0.005280		
Cluster_4700	11	11	9	0	2	12	05	144	Inf	glyoxalase
							6.73E-	0.005280		GNAT family
Cluster_4699	11	11	9	0	2	12	05	144	Inf	acetyltransferase
							6.73E-	0.005280		
Cluster_4725	11	11	9	0	2	12	05	144	Inf	histidine kinase
							6.73E-	0.005280		
Cluster_4805	11	11	9	0	2	12	05	144	Inf	histidine kinase
							6.73E-	0.005280		
Cluster_4806	11	11	9	0	2	12	05	144	Inf	histidine kinase
							6.73E-	0.005280		histidine phosphatase
Cluster_4729	11	11	9	0	2	12	05	144	Inf	family protein
							6.73E-	0.005280		
Cluster_4358	13	11	9	0	2	12	05	144	Inf	hypothetical protein
							6.73E-	0.005280		
Cluster_4697	11	11	9	0	2	12	05	144	Inf	hypothetical protein
							6.73E-	0.005280		
Cluster_4698	11	11	9	0	2	12	05	144	Inf	hypothetical protein
							6.73E-	0.005280		
Cluster_4704	11	11	9	0	2	12	05	144	Inf	hypothetical protein
							6.73E-	0.005280		
Cluster_4705	11	11	9	0	2	12	05	144	Inf	hypothetical protein
							6.73E-	0.005280		
Cluster_4707	11	11	9	0	2	12	05	144	Inf	hypothetical protein
							6.73E-	0.005280		
Cluster_4709	11	11	9	0	2	12	05	144	Inf	hypothetical protein
							6.73E-	0.005280		
Cluster_4721	11	11	9	0	2	12	05	144	Inf	hypothetical protein
Cluster_4728	11	11	9	0	2	12	6.73E-	0.005280	Inf	hypothetical protein

							05	144		
Cluster_4737	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4740	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4742	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4743	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4744	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4746	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4748	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4750	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4767	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4782	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4787	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4791	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4792	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4797	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4800	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4801	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4802	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	hypothetical protein
Cluster_4804	11	11	9	0	2	12	6.73E-	0.005280	Inf	hypothetical protein

							05	144		
							6.73E-	0.005280		
Cluster_4808	11	11	9	0	2	12	05	144	Inf	hypothetical protein
							1.22E-	0.005280		
Cluster_4973	10	10	9	0	1	13	05	144	Inf	hypothetical protein
							1.22E-	0.005280		
Cluster_5006	10	10	9	0	1	13	05	144	Inf	hypothetical protein
							1.22E-	0.005280		
Cluster_5007	10	10	9	0	1	13	05	144	Inf	hypothetical protein
							1.22E-	0.005280		
Cluster_5010	10	10	9	0	1	13	05	144	Inf	hypothetical protein
							6.73E-	0.005280		
Cluster_4691	11	11	9	0	2	12	05	144	Inf	invasion protein
							6.73E-	0.005280		
Cluster_4799	11	11	9	0	2	12	05	144	Inf	invasion protein
							6.73E-	0.005280		LacI family transcriptional
Cluster_4718	11	11	9	0	2	12	05	144	Inf	regulator
							6.73E-	0.005280		LysR family
Cluster_4770	11	11	9	0	2	12	05	144	Inf	transcriptional regulator
							6.73E-	0.005280		
Cluster_4738	11	11	9	0	2	12	05	144	Inf	MepB
							6.73E-	0.005280		
Cluster_4788	11	11	9	0	2	12	05	144	Inf	methyltransferase
							6.73E-	0.005280		PadR family
Cluster_4796	11	11	9	0	2	12	05	144	Inf	transcriptional regulator
							6.73E-	0.005280		
Cluster_4761	11	11	9	0	2	12	05	144	Inf	peptidase G2
							6.73E-	0.005280		
Cluster_4708	11	11	9	0	2	12	05	144	Inf	peptide transporter
							6.73E-	0.005280		
Cluster_4798	11	11	9	0	2	12	05	144	Inf	peptide-binding protein
							6.73E-	0.005280		
Cluster_4809	11	11	9	0	2	12	05	144	Inf	permease
							6.73E-	0.005280		
Cluster_4771	11	11	9	0	2	12	05	144	Inf	phage tail protein
Cluster_4807	11	11	9	0	2	12	6.73E-	0.005280	Inf	phosphoglycerol

							05	144		transferase
Cluster_4722	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	preprotein translocase
Cluster_4710	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	ROK family transcriptional regulator
Cluster_5008	10	10	9	0	1	13	1.22E-05	0.005280144	Inf	serine hydrolase
Cluster_4963	10	10	9	0	1	13	1.22E-05	0.005280144	Inf	serine/threonine protein kinase
Cluster_4702	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	siderophore biosynthesis protein
Cluster_4389	13	11	9	0	2	12	6.73E-05	0.005280144	Inf	sporulation protein
Cluster_4789	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	sporulation protein
Cluster_4832	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	sugar ABC transporter ATP-binding protein
Cluster_4692	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	teicoplanin resistance protein VanZ
Cluster_4943	10	10	9	0	1	13	1.22E-05	0.005280144	Inf	transglycosylase
Cluster_4367	13	11	9	0	2	12	6.73E-05	0.005280144	Inf	translation initiation inhibitor
Cluster_4706	11	11	9	0	2	12	6.73E-05	0.005280144	Inf	uridine kinase
Cluster_4510	12	12	9	0	3	11	0.0003	0.012648	Inf	3-phosphoshikimate 1-carboxyvinyltransferase
Cluster_4518	12	12	9	0	3	11	0.0003	0.012648	Inf	4-hydroxy-2-ketovalerate aldolase
Cluster_4533	12	12	9	0	3	11	0.0003	0.012648	Inf	ABC transporter
Cluster_4541	12	12	9	0	3	11	0.0003	0.012648	Inf	ABC transporter
Cluster_4523	12	12	9	0	3	11	0.0003	0.012648	Inf	alkaline phosphatase
Cluster_4449	12	12	9	0	3	11	0.0003	0.012648	Inf	alpha/beta hydrolase

							36519	25		
							0.0003	0.012648		betaine-aldehyde
Cluster_4507	12	12	9	0	3	11	36519	25	Inf	dehydrogenase
							0.0003	0.012648		damage-inducible protein
Cluster_4457	12	12	9	0	3	11	36519	25	Inf	DinB
							0.0003	0.012648		
Cluster_4511	12	12	9	0	3	11	36519	25	Inf	DNA helicase
							0.0003	0.012648		DNA-binding response
Cluster_4450	12	12	9	0	3	11	36519	25	Inf	regulator
							0.0003	0.012648		
Cluster_4529	12	12	9	0	3	11	36519	25	Inf	epimerase
							0.0003	0.012648		Fis family transcriptional
Cluster_4506	12	12	9	0	3	11	36519	25	Inf	regulator
							0.0003	0.012648		GntR family
Cluster_4526	12	12	9	0	3	11	36519	25	Inf	transcriptional regulator
							0.0003	0.012648		
Cluster_4432	12	12	9	0	3	11	36519	25	Inf	hypothetical protein
							0.0003	0.012648		
Cluster_4505	12	12	9	0	3	11	36519	25	Inf	hypothetical protein
							0.0003	0.012648		
Cluster_4509	12	12	9	0	3	11	36519	25	Inf	hypothetical protein
							0.0003	0.012648		
Cluster_4522	12	12	9	0	3	11	36519	25	Inf	hypothetical protein
							0.0003	0.012648		
Cluster_4537	12	12	9	0	3	11	36519	25	Inf	hypothetical protein
							0.0003	0.012648		MBL fold metallo-
Cluster_4530	12	12	9	0	3	11	36519	25	Inf	hydrolase
							0.0003	0.012648		
Cluster_4527	12	12	9	0	3	11	36519	25	Inf	MFS transporter
							0.0003	0.012648		
Cluster_4536	12	12	9	0	3	11	36519	25	Inf	MFS transporter
							0.0003	0.012648		
Cluster_4539	12	12	9	0	3	11	36519	25	Inf	MFS transporter
							0.0003	0.012648		multidrug ABC transporter
Cluster_4532	12	12	9	0	3	11	36519	25	Inf	ATP-binding protein
Cluster_4508	12	12	9	0	3	11	0.0003	0.012648	Inf	putrescine

							36519	25		aminotransferase
							0.0003	0.012648		
Cluster_4504	12	12	9	0	3	11	36519	25	Inf	Putrescine importer PuuP
							0.0003	0.012648		
Cluster_4535	12	12	9	0	3	11	36519	25	Inf	short-chain dehydrogenase
							0.0003	0.012648		
Cluster_4500	12	12	9	0	3	11	36519	25	Inf	sporulation protein
							0.0016	0.041052		ABC transporter ATP-
Cluster_4370	13	13	9	0	4	10	03054	31	Inf	binding protein
							0.0016	0.041052		
Cluster_4377	13	13	9	0	4	10	03054	31	Inf	amino acid permease
							0.0016	0.041052		cyclic nucleotide-binding
Cluster_4336	13	13	9	0	4	10	03054	31	Inf	protein
							0.0016	0.041052		cysteine ABC transporter
Cluster_4407	13	13	9	0	4	10	03054	31	Inf	permease
							0.0016	0.041052		DUF3948 domain-
Cluster_4213	15	13	9	0	4	10	03054	31	Inf	containing protein
							0.0016	0.041052		
Cluster_4379	13	13	9	0	4	10	03054	31	Inf	Fe-S oxidoreductase
							0.0016	0.041052		
Cluster_4380	13	13	9	0	4	10	03054	31	Inf	hydroxyglutarate oxidase
							0.0016	0.041052		
Cluster_4345	13	13	9	0	4	10	03054	31	Inf	hypothetical protein
							0.0016	0.041052		
Cluster_4375	13	13	9	0	4	10	03054	31	Inf	hypothetical protein
							0.0016	0.041052		
Cluster_4376	13	13	9	0	4	10	03054	31	Inf	hypothetical protein
							0.0016	0.041052		
Cluster_4388	13	13	9	0	4	10	03054	31	Inf	hypothetical protein
							0.0016	0.041052		
Cluster_4378	13	13	9	0	4	10	03054	31	Inf	lactate utilization protein C
							0.0016	0.041052		NADPH:quinone
Cluster_4363	13	13	9	0	4	10	03054	31	Inf	oxidoreductase
							0.0016	0.041052		
Cluster_4382	13	13	9	0	4	10	03054	31	Inf	ornithine cyclodeaminase
Cluster_4381	13	13	9	0	4	10	0.0016	0.041052	Inf	protein CsiD

							03054	31		
							0.0016	0.041052		ribonucleotide-diphosphate
Cluster_4409	13	13	9	0	4	10	03054	31	Inf	reductase subunit
							0.0016	0.041052		
Cluster_4371	13	13	9	0	4	10	03054	31	Inf	spore coat protein
							0.0016	0.041052		stage II sporulation protein
Cluster_4359	13	13	9	0	4	10	03054	31	Inf	P

[a] P-values from two-sided Fisher's Exact Tests

[b] P-values were corrected using the False Discovery Rate (FDR) approach

[c] Odds ratios marked as INF (Infinite) are a result of dividing by zero

Supplemental Table 4.3: OrthoMCL results: List of genes encoded in the genomes of *B. cereus* group isolates that cannot grow at 6°C

Cluster	Number of Genes	Number of Taxa	Presence Cold Growers	Absence Cold Growers	Presence Non-Cold Growers	Absence Non-Cold Growers	P-value [a]	FDR P-value [b]	Odds Ratio [c]	Products
Cluster_4415	12	12	0	9	12	2	6.73E-05	0.005280144	0	50S ribosomal protein L33
Cluster_4465	12	12	0	9	12	2	6.73E-05	0.005280144	0	ferrochelatase
Cluster_4453	12	12	0	9	12	2	6.73E-05	0.005280144	0	HAD family hydrolase
Cluster_4222	14	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4298	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4300	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4301	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4304	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4309	13	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4314	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4319	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4320	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4328	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4329	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4335	13	13	0	9	13	1	1.22E-	0.005280	0	hypothetical protein

							05	144		
Cluster_4339	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4350	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4355	13	13	0	9	13	1	1.22E-05	0.005280144	0	hypothetical protein
Cluster_4364	13	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4461	12	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4462	12	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4463	12	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4473	12	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4474	12	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4490	12	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4494	12	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4515	12	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4524	12	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4531	12	12	0	9	12	2	6.73E-05	0.005280144	0	hypothetical protein
Cluster_4413	12	12	0	9	12	2	6.73E-05	0.005280144	0	transporter
Cluster_4434	12	12	0	9	12	2	6.73E-05	0.005280144	0	ubiquinone biosynthesis methyltransferase UbiE
Cluster_4452	12	12	0	9	12	2	6.73E-05	0.005280144	0	XRE family transcriptional regulator
Cluster_4437	12	11	0	9	11	3	0.0003	0.012648	0	2'-5' RNA ligase

							36519	25		
							0.0003	0.012648		ABC transporter ATP-
Cluster_4590	11	11	0	9	11	3	36519	25	0	binding protein
							0.0003	0.012648		ABC transporter ATP-
Cluster_4615	11	11	0	9	11	3	36519	25	0	binding protein
							0.0003	0.012648		
Cluster_4477	12	11	0	9	11	3	36519	25	0	ABC transporter permease
							0.0003	0.012648		
Cluster_4649	11	11	0	9	11	3	36519	25	0	ABC transporter permease
							0.0003	0.012648		
Cluster_4624	11	11	0	9	11	3	36519	25	0	acetyl-CoA hydrolase
							0.0003	0.012648		
Cluster_4416	12	11	0	9	11	3	36519	25	0	acetyltransferases
							0.0003	0.012648		
Cluster_4571	11	11	0	9	11	3	36519	25	0	acetyltransferase
							0.0003	0.012648		
Cluster_4677	11	11	0	9	11	3	36519	25	0	acetyltransferase
							0.0003	0.012648		
Cluster_4683	11	11	0	9	11	3	36519	25	0	acetyltransferase
							0.0003	0.012648		
Cluster_4604	11	11	0	9	11	3	36519	25	0	alkaline phosphatase
							0.0003	0.012648		alkanesulfonate
Cluster_4438	12	11	0	9	11	3	36519	25	0	monooxygenase
							0.0003	0.012648		
Cluster_4657	11	11	0	9	11	3	36519	25	0	alpha/beta hydrolase
							0.0003	0.012648		
Cluster_4682	11	11	0	9	11	3	36519	25	0	alpha/beta hydrolase
							0.0003	0.012648		aminoglycoside
Cluster_4631	11	11	0	9	11	3	36519	25	0	phosphotransferase
							0.0003	0.012648		antibiotic biosynthesis
Cluster_4653	11	11	0	9	11	3	36519	25	0	monooxygenase
							0.0003	0.012648		AraC family
Cluster_4476	12	11	0	9	11	3	36519	25	0	transcriptional regulator
							0.0003	0.012648		AraC family
Cluster_4585	11	11	0	9	11	3	36519	25	0	transcriptional regulator
Cluster_4628	11	11	0	9	11	3	0.0003	0.012648	0	cell surface protein"

							36519	25		
							0.0003	0.012648		chromosome segregation
Cluster_4650	11	11	0	9	11	3	36519	25	0	protein
							0.0003	0.012648		
Cluster_4578	11	11	0	9	11	3	36519	25	0	collagenase
							0.0003	0.012648		
Cluster_4619	11	11	0	9	11	3	36519	25	0	competence protein ComF
							0.0003	0.012648		
Cluster_4573	11	11	0	9	11	3	36519	25	0	competence protein ComG
							0.0003	0.012648		
Cluster_4574	11	11	0	9	11	3	36519	25	0	competence protein ComG
							0.0003	0.012648		
Cluster_4575	11	11	0	9	11	3	36519	25	0	competence protein ComG
							0.0003	0.012648		
Cluster_4576	11	11	0	9	11	3	36519	25	0	competence protein
							0.0003	0.012648		
Cluster_4618	11	11	0	9	11	3	36519	25	0	competence protein
							0.0003	0.012648		
Cluster_4629	11	11	0	9	11	3	36519	25	0	cytochrome B
							0.0003	0.012648		delta-aminolevulinic acid
Cluster_4570	11	11	0	9	11	3	36519	25	0	dehydratase
							0.0003	0.012648		DNA mismatch repair
Cluster_4642	11	11	0	9	11	3	36519	25	0	protein MutT
							0.0003	0.012648		
Cluster_4658	11	11	0	9	11	3	36519	25	0	DNA-binding protein
							0.0003	0.012648		glycerophosphodiester
Cluster_4640	11	11	0	9	11	3	36519	25	0	phosphodiesterase
							0.0003	0.012648		glycosyl transferase family
Cluster_4622	11	11	0	9	11	3	36519	25	0	2
							0.0003	0.012648		
Cluster_4659	11	11	0	9	11	3	36519	25	0	glycosyl transferase
							0.0003	0.012648		
Cluster_4572	11	11	0	9	11	3	36519	25	0	group-specific protein
							0.0003	0.012648		
Cluster_4644	11	11	0	9	11	3	36519	25	0	group-specific protein
Cluster_4564	11	11	0	9	11	3	0.0003	0.012648	0	histidine kinase

							36519	25		
							0.0003	0.012648		
Cluster_4652	11	11	0	9	11	3	36519	25	0	histidine kinase
							0.0003	0.012648		
Cluster_4669	11	11	0	9	11	3	36519	25	0	histidine kinase
							0.0003	0.012648		histidine phosphatase
Cluster_4569	11	11	0	9	11	3	36519	25	0	family protein
							0.0003	0.012648		
Cluster_4436	12	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4566	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4568	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4592	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4596	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4598	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4602	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4610	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4614	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4623	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4627	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4630	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4632	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4633	11	11	0	9	11	3	36519	25	0	hypothetical protein
Cluster_4641	11	11	0	9	11	3	0.0003	0.012648	0	hypothetical protein

							36519	25		
							0.0003	0.012648		
Cluster_4643	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4651	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4654	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4655	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4656	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4671	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4680	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4688	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4712	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4716	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4720	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4724	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4752	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4753	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4758	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4760	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4765	11	11	0	9	11	3	36519	25	0	hypothetical protein
Cluster_4783	11	11	0	9	11	3	0.0003	0.012648	0	hypothetical protein

							36519	25		
							0.0003	0.012648		
Cluster_4810	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		
Cluster_4815	11	11	0	9	11	3	36519	25	0	hypothetical protein
							0.0003	0.012648		LacI family transcriptional
Cluster_4417	12	11	0	9	11	3	36519	25	0	regulator
							0.0003	0.012648		LytR family transcriptional
Cluster_4625	11	11	0	9	11	3	36519	25	0	regulator
							0.0003	0.012648		
Cluster_4608	11	11	0	9	11	3	36519	25	0	magnesium transporter
							0.0003	0.012648		
Cluster_4595	11	11	0	9	11	3	36519	25	0	mep operon protein MepB
							0.0003	0.012648		
Cluster_4679	11	11	0	9	11	3	36519	25	0	metal-dependent hydrolase
							0.0003	0.012648		
Cluster_4587	11	11	0	9	11	3	36519	25	0	methyltransferase
							0.0003	0.012648		
Cluster_4588	11	11	0	9	11	3	36519	25	0	methyltransferase
							0.0003	0.012648		
Cluster_4597	11	11	0	9	11	3	36519	25	0	MFS transporter
							0.0003	0.012648		molybdopterin synthase
Cluster_4645	11	11	0	9	11	3	36519	25	0	sulfur carrier subunit
							0.0003	0.012648		
Cluster_4635	11	11	0	9	11	3	36519	25	0	peptidase G2
							0.0003	0.012648		phenazine biosynthesis
Cluster_4607	11	11	0	9	11	3	36519	25	0	protein PhzF
							0.0003	0.012648		
Cluster_4670	11	11	0	9	11	3	36519	25	0	rhodanese
							0.0003	0.012648		RNA polymerase subunit
Cluster_4617	11	11	0	9	11	3	36519	25	0	sigma-70
							0.0003	0.012648		serine/threonine protein
Cluster_4591	11	11	0	9	11	3	36519	25	0	kinase
							0.0003	0.012648		
Cluster_4647	11	11	0	9	11	3	36519	25	0	spore gernimation protein
Cluster_4426	12	11	0	9	11	3	0.0003	0.012648	0	sporulation protein

							36519	25		
							0.0003	0.012648		stage II sporulation protein
Cluster_4621	11	11	0	9	11	3	36519	25	0	E
							0.0003	0.012648		sugar ABC transporter
Cluster_4603	11	11	0	9	11	3	36519	25	0	ATP-binding protein
							0.0003	0.012648		sulfonate ABC transporter
Cluster_4439	12	11	0	9	11	3	36519	25	0	permease
							0.0003	0.012648		sulfonate ABC transporter
Cluster_4440	12	11	0	9	11	3	36519	25	0	substrate-binding
							0.0003	0.012648		translation initiation
Cluster_4589	11	11	0	9	11	3	36519	25	0	inhibitor
							0.0003	0.012648		
Cluster_4418	12	11	0	9	11	3	36519	25	0	transporter
							0.0003	0.012648		two-component sensor
Cluster_4563	11	11	0	9	11	3	36519	25	0	histidine kinase
							0.0003	0.012648		
Cluster_4605	11	11	0	9	11	3	36519	25	0	uridine kinase
							0.0003	0.012648		YlmC/YmxH family
Cluster_4577	11	11	0	9	11	3	36519	25	0	sporulation protein
							0.0016	0.041052		
Cluster_4594	11	10	0	9	10	4	03054	31	0	acetyltransferase
							0.0016	0.041052		
Cluster_4601	11	10	0	9	10	4	03054	31	0	acetyltransferase
							0.0016	0.041052		
Cluster_4878	10	10	0	9	10	4	03054	31	0	acetyltransferase
							0.0016	0.041052		
Cluster_4899	10	10	0	9	10	4	03054	31	0	alkaline serine protease
							0.0016	0.041052		aminoglycoside
Cluster_4992	10	10	0	9	10	4	03054	31	0	phosphotransferase
							0.0016	0.041052		
Cluster_4896	10	10	0	9	10	4	03054	31	0	ATP-binding protein
							0.0016	0.041052		
Cluster_4835	10	10	0	9	10	4	03054	31	0	cell division protein FtsN
							0.0016	0.041052		
Cluster_4870	10	10	0	9	10	4	03054	31	0	cell wall anchor protein
Cluster_4995	10	10	0	9	10	4	0.0016	0.041052	0	DNA recombination

							03054	31		protein RecO
							0.0016	0.041052		
Cluster_4869	10	10	0	9	10	4	03054	31	0	group-specific protein
							0.0016	0.041052		
Cluster_4841	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4843	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4852	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4862	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4865	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4874	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4889	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4891	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4898	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4904	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4905	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4912	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4935	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4982	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4994	10	10	0	9	10	4	03054	31	0	hypothetical protein
							0.0016	0.041052		
Cluster_4997	10	10	0	9	10	4	03054	31	0	hypothetical protein
Cluster_4998	10	10	0	9	10	4	0.0016	0.041052	0	hypothetical protein

							03054	31		
							0.0016	0.041052		MerR family
Cluster_4890	10	10	0	9	10	4	03054	31	0	transcriptional regulator
							0.0016	0.041052		
Cluster_4987	10	10	0	9	10	4	03054	31	0	MFS transporter
							0.0016	0.041052		N-acetylmuramoyl-L-
Cluster_4868	10	10	0	9	10	4	03054	31	0	alanine amidase
							0.0016	0.041052		
Cluster_4855	10	10	0	9	10	4	03054	31	0	oligoendopeptidase F
							0.0016	0.041052		
Cluster_4834	10	10	0	9	10	4	03054	31	0	peptidase S8
							0.0016	0.041052		
Cluster_4856	10	10	0	9	10	4	03054	31	0	peptide-binding protein
							0.0016	0.041052		RpiR family
Cluster_4981	10	10	0	9	10	4	03054	31	0	transcriptional regulator
							0.0016	0.041052		siderophore biosynthesis
Cluster_4908	10	10	0	9	10	4	03054	31	0	protein
							0.0016	0.041052		
Cluster_4895	10	10	0	9	10	4	03054	31	0	spore gernimation protein
							0.0016	0.041052		
Cluster_4599	11	10	0	9	10	4	03054	31	0	sporulation protein
							0.0016	0.041052		
Cluster_4600	11	10	0	9	10	4	03054	31	0	sporulation protein
							0.0016	0.041052		
Cluster_4897	10	10	0	9	10	4	03054	31	0	sugar kinase
							0.0016	0.041052		sulfonate ABC transporter
Cluster_4857	10	10	0	9	10	4	03054	31	0	ATP-binding protein
							0.0016	0.041052		
Cluster_4864	10	10	0	9	10	4	03054	31	0	transcriptional regulator
									0.027	
							0.0007	0.023753	61605	DNA-binding response
Cluster_4352	13	13	1	8	12	2	28105	52	1	regulator
									0.027	
							0.0007	0.023753	61605	
Cluster_4312	13	13	1	8	12	2	28105	52	1	general stress protein
Cluster_4308	13	13	1	8	12	2	0.0007	0.023753	0.027	hypothetical protein

							28105	52	616051	
									0.027	
							0.0007	0.023753	616051	
Cluster_4332	13	13	1	8	12	2	28105	52		hypothetical protein
									0.027	
							0.0007	0.023753	616051	
Cluster_4342	13	13	1	8	12	2	28105	52		hypothetical protein
									0.027	
							0.0007	0.023753	616051	
Cluster_4353	13	13	1	8	12	2	28105	52		hypothetical protein
									0.027	
							0.0007	0.023753	616051	two-component sensor
Cluster_4351	13	13	1	8	12	2	28105	52		histidine kinase
									0.029	
							0.0010	0.031628	074128	
Cluster_4158	15	15	2	7	13	1	46268	02		CAAX protease

[a] P-values from two-sided Fisher's Exact Tests

[b] P-values were corrected using the False Discovery Rate (FDR) approach

[c] Odds ratios marked as INF (Infinite) are a result of dividing by zero

Supplemental Table 4.4: List of Gene Ontology terms significantly overrepresented in the genomes of *B. cereus* group isolates that can grow at 6°C

GO Terms	Presence Cold Growers	Absence Cold Growers	Presence Non-Cold Growers	Absence Non-Cold Growers	p-values	Odds Ratio	FDR Corrected p-value	Description	GO Term Category	EC Annotation
GO:0005886	5032	1340	7576	2336	1.55E-04	1.157882741	0.03291156	plasma membrane	cellular component	NA[a]
GO:0005524	4288	1535	6353	2705	3.92E-06	1.189403566	0.004796152	ATP binding	molecular function	NA[a]
GO:0016020	3193	1370	4724	2374	1.13E-04	1.171231593	0.02989232	membrane	cellular component	NA[a]
GO:0006810	1377	801	1966	1422	1.22E-04	1.243336525	0.02989232	transport	biological process	NA[a]
GO:0005622	1103	490	1515	963	1.39E-07	1.430708714	0.000340728	intracellular	cellular component	NA[a]
GO:0009365	628	272	858	542	3.23E-05	1.458236169	0.01972535	protein histidine kinase complex	cellular component	NA[a]
GO:0000155	612	270	832	540	2.35E-05	1.470888677	0.01639143	phosphorelay sensor kinase activity	molecular function	NA[a]
GO:0016616	147	69	174	162	1.98E-04	1.981062734	0.03291156	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	molecular function	NA[a]
GO:0055088	86	4	106	34	4.18E-05	6.849898856	0.022691	lipid homeostasis	biological process	NA[a]
GO:0009447	36	9	24	46	1.49E-06	7.515781593	0.002422983	putrescine catabolic process	biological process	NA[a]
EC:3.6.3.25	35	1	32	24	8.75E-06	25.55086627	0.007132585	sulfate-transporting ATPase	NA	NA[a]

GO:0015419	35	1	32	24	8.75E-06	25.55086627	0.007132585	ATPase-coupled sulfate transmembrane transporter activity	molecular function	NA[a]
GO:0009263	27	0	26	16	9.64E-05	NA[a]	0.02774262	deoxyribonucleotide biosynthetic process	biological process	NA[a]
GO:0005768	26	10	18	38	2.45E-04	5.376904943	0.03291156	endosome	cellular component	NA[a]
EC:1.2.1.4	25	2	20	22	9.51E-05	13.25817367	0.02774262	Aldehyde dehydrogenase (NADP(+))	NA[a]	NA[a]
GO:0033721	25	2	20	22	9.51E-05	13.25817367	0.02774262	aldehyde dehydrogenase (NADP+) activity	molecular function	NA[a]
GO:0051063	18	0	12	16	6.02E-05	NA[a]	0.02350826	CDP reductase activity	molecular function	NA[a]
EC:3.1.1.31	9	0	2	12	6.73E-05	NA[a]	0.02350826	6-phospho gluconolactonase.	NA[a]	NA[a]
GO:0008655	9	0	2	12	6.73E-05	NA[a]	0.02350826	pyrimidine-containing compound salvage	biological process	NA[a]
GO:0009174	9	0	2	12	6.73E-05	NA[a]	0.02350826	pyrimidine ribonucleoside monophosphate biosynthetic process	biological process	NA[a]
GO:0017057	9	0	2	12	6.73E-05	NA[a]	0.02350826	6-phosphogluconolactonase activity	molecular function	NA[a]
EC:2.6.1.82	9	0	3	11	3.37E-04	NA[a]	0.03291156	PAT. Putrescine transaminase Putrescine aminotransferase.	NA[a]	Putrescine-alpha-ketoglutarate

										transaminase
EC:3.7.1.14	9	0	3	11	3.37E-04	NA[a]	0.03291156	2-hydroxy-6-oxonona-2,4-dienedioate hydrolase.	NA[a]	NA[a]
GO:0006082	9	0	3	11	3.37E-04	NA[a]	0.03291156	organic acid metabolic process	biological process	NA[a]
GO:0009258	9	0	3	11	3.37E-04	NA[a]	0.03291156	10-formyltetrahydrofolate catabolic process	biological process	NA[a]
GO:0015489	9	0	3	11	3.37E-04	NA[a]	0.03291156	putrescine transmembrane transporter activity	molecular function	NA[a]
GO:0018771	9	0	3	11	3.37E-04	NA[a]	0.03291156	2-hydroxy-6-oxonona-2,4-dienedioate hydrolase activity	molecular function	NA[a]
GO:0032383	9	0	3	11	3.37E-04	NA[a]	0.03291156	regulation of intracellular cholesterol transport	biological process	NA[a]
GO:0033094	9	0	3	11	3.37E-04	NA[a]	0.03291156	butane-1,4-diamine:2-oxoglutarate aminotransferase activity	molecular function	NA[a]
GO:0042632	9	0	3	11	3.37E-04	NA[a]	0.03291156	cholesterol homeostasis	biological process	NA[a]
GO:0048545	9	0	3	11	3.37E-04	NA[a]	0.03291156	response to steroid hormone	biological process	NA[a]
GO:0048870	9	0	3	11	3.37E-04	NA[a]	0.03291156	cell motility	biological process	NA[a]
GO:0052823	9	0	3	11	3.37E-04	NA[a]	0.03291156	2-hydroxy-6-oxonona-2,4,7-trienedioate	molecular function	NA[a]

								hydrolase activity		
GO:00									biological	
55081	9	0	3	11	3.37E-04	NA[a]	0.03291156	anion homeostasis	process	NA[a]

[a] NA: Not Applicable

Supplemental Table 4.5: List of Gene Ontology terms significantly overrepresented in the genomes of *B. cereus* group isolates that cannot grow at 6°C

GO Terms	Presence Cold Growers	Absence Cold Growers	Presence Non-Cold Growers	Absence Non-Cold Growers	p-values	Odds Ratio	FDR Corrected p-value	Description	GO Term Category	EC Annotation
GO:0042959	0	18	22	6	5.03E-08	0	0.000246171	alkanesulfonate transporter activity	molecular function	NA
EC:1.14.14.5	0	9	11	3	3.37E-04	0	0.03291156	alkanesulfonate monooxygenase	NA[a]	FMNH(2)-dependent aliphatic sulfonate monooxygenase. Sulfate starvation-induced protein 6.
GO:0006084	0	9	11	3	3.37E-04	0	0.03291156	acetyl-CoA metabolic process	biological process	NA[a]
GO:0006836	0	9	11	3	3.37E-04	0	0.03291156	neurotransmitter transport	biological process	NA[a]
GO:0007420	0	9	11	3	3.37E-04	0	0.03291156	brain development	biological process	NA[a]
GO:0008504	0	9	11	3	3.37E-04	0	0.03291156	monoamine transmembrane transporter activity	molecular function	NA[a]
GO:0008726	0	9	11	3	3.37E-04	0	0.03291156	alkanesulfonate monooxygenase activity	molecular function	NA[a]
GO:0015370	0	9	11	3	3.37E-04	0	0.03291156	solute:sodium symporter activity	molecular function	NA[a]

GO:00 15807	0	9	11	3	3.37E-04	0	0.03291156	L-amino acid transport	biological process	NA[a]
GO:00 19427	0	9	11	3	3.37E-04	0	0.03291156	acetyl-CoA biosynthetic process from acetate	biological process	NA[a]
GO:00 46306	0	9	11	3	3.37E-04	0	0.03291156	alkanesulfonate catabolic process	biological process	NA[a]
GO:00 51715	0	18	14	14	1.98E-04	0	0.03291156	cytolysis in other organism	biological process	NA[a]
GO:00 90101	0	9	11	3	3.37E-04	0	0.03291156	negative regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	biological process	NA[a]
GO:19 01289	0	9	11	3	3.37E-04	0	0.03291156	succinyl-CoA catabolic process	biological process	NA[a]
GO:00 10038	1	26	19	23	1.22E-04	0.048	0.02989232	response to metal ion	biological process	NA[a]

Supplemental Table 4.6: Hidden Markov Model (HMM) results for 23 *B. cereus* group isolates

Query	Accession	Strain	Count Match Per Genome[a]	Count Genomes[b]
Caps_synth_CapC	PF14102.5	FSL_E2-0214	2	1
Caps_synth_CapC	PF14102.5	FSL_H7-0683	2	2
Caps_synth_CapC	PF14102.5	FSL_H7-0926	2	3
Caps_synth_CapC	PF14102.5	FSL_H8-0485	2	4
Caps_synth_CapC	PF14102.5	FSL_H8-0492	2	5
Caps_synth_CapC	PF14102.5	FSL_H8-0534	1	6
Caps_synth_CapC	PF14102.5	FSL_J3-0113	2	7
Caps_synth_CapC	PF14102.5	FSL_K6-0069	3	8
Caps_synth_CapC	PF14102.5	FSL_K6-1030	1	9
Caps_synth_CapC	PF14102.5	FSL_M7-0109	2	10
Caps_synth_CapC	PF14102.5	FSL_M7-0669	2	11
Caps_synth_CapC	PF14102.5	FSL_M7-1219	1	12
Caps_synth_CapC	PF14102.5	FSL_M8-0091	1	13
Caps_synth_CapC	PF14102.5	FSL_M8-0117	1	14
Caps_synth_CapC	PF14102.5	FSL_M8-0473	2	15
Caps_synth_CapC	PF14102.5	FSL_R5-0708	3	16
Caps_synth_CapC	PF14102.5	FSL_R5-0811	3	17
Caps_synth_CapC	PF14102.5	FSL_W7-1108	1	18
Caps_synth_CapC	PF14102.5	FSL_W8-0050	2	19
Caps_synth_CapC	PF14102.5	FSL_W8-0169	2	20
Caps_synth_CapC	PF14102.5	FSL_W8-0268	3	21
Caps_synth_CapC	PF14102.5	FSL_W8-0483	2	22
CSD	PF00313.21	FSL_E2-0214	7	1
CSD	PF00313.21	FSL_H7-0683	6	2
CSD	PF00313.21	FSL_H7-0926	7	3
CSD	PF00313.21	FSL_H8-0485	7	4
CSD	PF00313.21	FSL_H8-0492	6	5
CSD	PF00313.21	FSL_H8-0534	6	6
CSD	PF00313.21	FSL_J3-0113	6	7
CSD	PF00313.21	FSL_J3-0123	7	8
CSD	PF00313.21	FSL_K6-0069	6	9
CSD	PF00313.21	FSL_K6-1030	6	10
CSD	PF00313.21	FSL_M7-0109	7	11
CSD	PF00313.21	FSL_M7-0669	6	12
CSD	PF00313.21	FSL_M7-1219	7	13
CSD	PF00313.21	FSL_M8-0091	6	14
CSD	PF00313.21	FSL_M8-0117	6	15
CSD	PF00313.21	FSL_M8-0473	6	16
CSD	PF00313.21	FSL_R5-0708	7	17

CSD	PF00313.21	FSL_R5-0811	6	18
CSD	PF00313.21	FSL_W7-1108	7	19
CSD	PF00313.21	FSL_W8-0050	7	20
CSD	PF00313.21	FSL_W8-0169	7	21
CSD	PF00313.21	FSL_W8-0268	6	22
CSD	PF00313.21	FSL_W8-0483	6	23
DEAD	PF00270.28	FSL_E2-0214	36	1
DEAD	PF00270.28	FSL_H7-0683	33	2
DEAD	PF00270.28	FSL_H7-0926	35	3
DEAD	PF00270.28	FSL_H8-0485	37	4
DEAD	PF00270.28	FSL_H8-0492	37	5
DEAD	PF00270.28	FSL_H8-0534	36	6
DEAD	PF00270.28	FSL_J3-0113	39	7
DEAD	PF00270.28	FSL_J3-0123	35	8
DEAD	PF00270.28	FSL_K6-0069	41	9
DEAD	PF00270.28	FSL_K6-1030	39	10
DEAD	PF00270.28	FSL_M7-0109	39	11
DEAD	PF00270.28	FSL_M7-0669	34	12
DEAD	PF00270.28	FSL_M7-1219	35	13
DEAD	PF00270.28	FSL_M8-0091	36	14
DEAD	PF00270.28	FSL_M8-0117	39	15
DEAD	PF00270.28	FSL_M8-0473	35	16
DEAD	PF00270.28	FSL_R5-0708	34	17
DEAD	PF00270.28	FSL_R5-0811	38	18
DEAD	PF00270.28	FSL_W7-1108	36	19
DEAD	PF00270.28	FSL_W8-0050	40	20
DEAD	PF00270.28	FSL_W8-0169	37	21
DEAD	PF00270.28	FSL_W8-0268	36	22
DEAD	PF00270.28	FSL_W8-0483	41	23
DnaJ	PF00226.30	FSL_E2-0214	1	1
DnaJ	PF00226.30	FSL_H7-0683	2	2
DnaJ	PF00226.30	FSL_H7-0926	1	3
DnaJ	PF00226.30	FSL_H8-0485	3	4
DnaJ	PF00226.30	FSL_H8-0492	1	5
DnaJ	PF00226.30	FSL_H8-0534	4	6
DnaJ	PF00226.30	FSL_J3-0113	2	7
DnaJ	PF00226.30	FSL_J3-0123	2	8
DnaJ	PF00226.30	FSL_K6-0069	3	9
DnaJ	PF00226.30	FSL_K6-1030	2	10
DnaJ	PF00226.30	FSL_M7-0109	1	11
DnaJ	PF00226.30	FSL_M7-0669	1	12
DnaJ	PF00226.30	FSL_M7-1219	2	13

DnaJ	PF00226.30	FSL_M8-0091	2	14
DnaJ	PF00226.30	FSL_M8-0117	1	15
DnaJ	PF00226.30	FSL_M8-0473	1	16
DnaJ	PF00226.30	FSL_R5-0708	1	17
DnaJ	PF00226.30	FSL_R5-0811	1	18
DnaJ	PF00226.30	FSL_W7-1108	2	19
DnaJ	PF00226.30	FSL_W8-0050	2	20
DnaJ	PF00226.30	FSL_W8-0169	2	21
DnaJ	PF00226.30	FSL_W8-0268	1	22
DnaJ	PF00226.30	FSL_W8-0483	3	23
FA_desaturase	PF00487.23	FSL_E2-0214	7	1
FA_desaturase	PF00487.23	FSL_H7-0683	7	2
FA_desaturase	PF00487.23	FSL_H7-0926	7	3
FA_desaturase	PF00487.23	FSL_H8-0485	10	4
FA_desaturase	PF00487.23	FSL_H8-0492	8	5
FA_desaturase	PF00487.23	FSL_H8-0534	8	6
FA_desaturase	PF00487.23	FSL_J3-0113	8	7
FA_desaturase	PF00487.23	FSL_J3-0123	7	8
FA_desaturase	PF00487.23	FSL_K6-0069	9	9
FA_desaturase	PF00487.23	FSL_K6-1030	9	10
FA_desaturase	PF00487.23	FSL_M7-0109	8	11
FA_desaturase	PF00487.23	FSL_M7-0669	6	12
FA_desaturase	PF00487.23	FSL_M7-1219	5	13
FA_desaturase	PF00487.23	FSL_M8-0091	9	14
FA_desaturase	PF00487.23	FSL_M8-0117	7	15
FA_desaturase	PF00487.23	FSL_M8-0473	6	16
FA_desaturase	PF00487.23	FSL_R5-0708	6	17
FA_desaturase	PF00487.23	FSL_R5-0811	6	18
FA_desaturase	PF00487.23	FSL_W7-1108	8	19
FA_desaturase	PF00487.23	FSL_W8-0050	9	20
FA_desaturase	PF00487.23	FSL_W8-0169	6	21
FA_desaturase	PF00487.23	FSL_W8-0268	8	22
FA_desaturase	PF00487.23	FSL_W8-0483	6	23
FA_desaturase_2	PF03405.13	FSL_E2-0214	1	1
FA_desaturase_2	PF03405.13	FSL_H7-0683	1	2
FA_desaturase_2	PF03405.13	FSL_H7-0926	2	3
FA_desaturase_2	PF03405.13	FSL_H8-0485	3	4
FA_desaturase_2	PF03405.13	FSL_H8-0492	1	5
FA_desaturase_2	PF03405.13	FSL_H8-0534	2	6
FA_desaturase_2	PF03405.13	FSL_J3-0113	1	7
FA_desaturase_2	PF03405.13	FSL_K6-0069	2	8
FA_desaturase_2	PF03405.13	FSL_K6-1030	3	9

FA_desaturase_2	PF03405.13	FSL_M7-0109	1	10
FA_desaturase_2	PF03405.13	FSL_M7-0669	2	11
FA_desaturase_2	PF03405.13	FSL_M7-1219	2	12
FA_desaturase_2	PF03405.13	FSL_M8-0091	1	13
FA_desaturase_2	PF03405.13	FSL_M8-0473	2	14
FA_desaturase_2	PF03405.13	FSL_R5-0708	1	15
FA_desaturase_2	PF03405.13	FSL_R5-0811	1	16
FA_desaturase_2	PF03405.13	FSL_W7-1108	4	17
FA_desaturase_2	PF03405.13	FSL_W8-0050	2	18
FA_desaturase_2	PF03405.13	FSL_W8-0169	2	19
FA_desaturase_2	PF03405.13	FSL_W8-0268	1	20
FA_hydroxylase	PF04116.12	FSL_E2-0214	2	1
FA_hydroxylase	PF04116.12	FSL_H7-0683	3	2
FA_hydroxylase	PF04116.12	FSL_H7-0926	2	3
FA_hydroxylase	PF04116.12	FSL_H8-0485	2	4
FA_hydroxylase	PF04116.12	FSL_H8-0492	2	5
FA_hydroxylase	PF04116.12	FSL_H8-0534	3	6
FA_hydroxylase	PF04116.12	FSL_J3-0113	1	7
FA_hydroxylase	PF04116.12	FSL_J3-0123	2	8
FA_hydroxylase	PF04116.12	FSL_K6-0069	1	9
FA_hydroxylase	PF04116.12	FSL_K6-1030	1	10
FA_hydroxylase	PF04116.12	FSL_M7-0109	2	11
FA_hydroxylase	PF04116.12	FSL_M7-0669	2	12
FA_hydroxylase	PF04116.12	FSL_M7-1219	1	13
FA_hydroxylase	PF04116.12	FSL_M8-0091	2	14
FA_hydroxylase	PF04116.12	FSL_M8-0117	1	15
FA_hydroxylase	PF04116.12	FSL_M8-0473	1	16
FA_hydroxylase	PF04116.12	FSL_R5-0708	3	17
FA_hydroxylase	PF04116.12	FSL_R5-0811	1	18
FA_hydroxylase	PF04116.12	FSL_W7-1108	3	19
FA_hydroxylase	PF04116.12	FSL_W8-0050	1	20
FA_hydroxylase	PF04116.12	FSL_W8-0169	2	21
FA_hydroxylase	PF04116.12	FSL_W8-0268	1	22
FA_hydroxylase	PF04116.12	FSL_W8-0483	1	23
LtrA	PF06772.10	FSL_M8-0473	1	1
LtrA	PF06772.10	FSL_R5-0811	1	2
Peptidase_S11	PF00768.19	FSL_E2-0214	11	1
Peptidase_S11	PF00768.19	FSL_H7-0683	11	2
Peptidase_S11	PF00768.19	FSL_H7-0926	13	3
Peptidase_S11	PF00768.19	FSL_H8-0485	13	4
Peptidase_S11	PF00768.19	FSL_H8-0492	12	5
Peptidase_S11	PF00768.19	FSL_H8-0534	13	6

Peptidase_S11	PF00768.19	FSL_J3-0113	12	7
Peptidase_S11	PF00768.19	FSL_J3-0123	16	8
Peptidase_S11	PF00768.19	FSL_K6-0069	14	9
Peptidase_S11	PF00768.19	FSL_K6-1030	13	10
Peptidase_S11	PF00768.19	FSL_M7-0109	13	11
Peptidase_S11	PF00768.19	FSL_M7-0669	11	12
Peptidase_S11	PF00768.19	FSL_M7-1219	16	13
Peptidase_S11	PF00768.19	FSL_M8-0091	12	14
Peptidase_S11	PF00768.19	FSL_M8-0117	11	15
Peptidase_S11	PF00768.19	FSL_M8-0473	11	16
Peptidase_S11	PF00768.19	FSL_R5-0708	11	17
Peptidase_S11	PF00768.19	FSL_R5-0811	12	18
Peptidase_S11	PF00768.19	FSL_W7-1108	16	19
Peptidase_S11	PF00768.19	FSL_W8-0050	13	20
Peptidase_S11	PF00768.19	FSL_W8-0169	13	21
Peptidase_S11	PF00768.19	FSL_W8-0268	11	22
Peptidase_S11	PF00768.19	FSL_W8-0483	15	23
RecA	PF00154.20	FSL_E2-0214	3	1
RecA	PF00154.20	FSL_H7-0683	3	2
RecA	PF00154.20	FSL_H7-0926	3	3
RecA	PF00154.20	FSL_H8-0485	3	4
RecA	PF00154.20	FSL_H8-0492	3	5
RecA	PF00154.20	FSL_H8-0534	6	6
RecA	PF00154.20	FSL_J3-0113	2	7
RecA	PF00154.20	FSL_J3-0123	3	8
RecA	PF00154.20	FSL_K6-0069	3	9
RecA	PF00154.20	FSL_K6-1030	4	10
RecA	PF00154.20	FSL_M7-0109	3	11
RecA	PF00154.20	FSL_M7-0669	5	12
RecA	PF00154.20	FSL_M7-1219	3	13
RecA	PF00154.20	FSL_M8-0091	4	14
RecA	PF00154.20	FSL_M8-0117	1	15
RecA	PF00154.20	FSL_M8-0473	4	16
RecA	PF00154.20	FSL_R5-0708	3	17
RecA	PF00154.20	FSL_R5-0811	4	18
RecA	PF00154.20	FSL_W7-1108	3	19
RecA	PF00154.20	FSL_W8-0050	3	20
RecA	PF00154.20	FSL_W8-0169	3	21
RecA	PF00154.20	FSL_W8-0268	3	22
RecA	PF00154.20	FSL_W8-0483	2	23
YdjO	PF14169.5	FSL_E2-0214	3	1
YdjO	PF14169.5	FSL_H7-0683	3	2

YdjO	PF14169.5	FSL_H7-0926	3	3
YdjO	PF14169.5	FSL_H8-0485	3	4
YdjO	PF14169.5	FSL_H8-0492	3	5
YdjO	PF14169.5	FSL_H8-0534	2	6
YdjO	PF14169.5	FSL_J3-0113	2	7
YdjO	PF14169.5	FSL_J3-0123	2	8
YdjO	PF14169.5	FSL_K6-0069	3	9
YdjO	PF14169.5	FSL_K6-1030	1	10
YdjO	PF14169.5	FSL_M7-0109	3	11
YdjO	PF14169.5	FSL_M7-0669	3	12
YdjO	PF14169.5	FSL_M7-1219	2	13
YdjO	PF14169.5	FSL_M8-0091	2	14
YdjO	PF14169.5	FSL_M8-0117	1	15
YdjO	PF14169.5	FSL_M8-0473	1	16
YdjO	PF14169.5	FSL_R5-0708	3	17
YdjO	PF14169.5	FSL_R5-0811	1	18
YdjO	PF14169.5	FSL_W7-1108	2	19
YdjO	PF14169.5	FSL_W8-0050	1	20
YdjO	PF14169.5	FSL_W8-0169	2	21
YdjO	PF14169.5	FSL_W8-0268	2	22
YdjO	PF14169.5	FSL_W8-0483	1	23

[a] Count Match Per Genome is indicative of the number of times a domain is found in a given genome

[b] Count Genomes is the running total of genomes in which a given domain query is found